THE UNIVERSITY OF ALBERTA

TESTING HYPOTHESES CONCERNING PARAMETERS WHICH ARE RANDOM VARIABLES

by

© HIMATLAL P. JAGANI

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

EDMONTON, ALBERTA

FALL, 1970

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "TESTING HYPOTHESES CONCERNING PARAMETERS WHICH ARE RANDOM VARIABLES", submitted by HIMATLAL P. JAGANI, in partial fulfilment of the requirements for the degree of Master of Science.

ABSTRACT

In Chapter I some developments in testing when the parameter, λ , is a random variable are briefly mentioned. The "compound decision problem" investigated by H. Robbins and "An empirical Bayes approach" considered by E. Samuel are dealt with in some detail. Mention is made of works by J. Hannan and G. Meeden.

In Chapter II Meeden's results, as embodied in Chapter III of his doctoral thesis, are studied in detail. Some results are stated and comments are made on his main theorems. In the process it is possible to present much of his work, more concisely, and without one main theorem.

In Chapter III a slightly different approach, based on the concept of a least favourable distribution, is taken to investigate the same problem. This results in a characterization of the least favourable distribution. Two corollaries are then proved. The construction of the UMP Bayesian test when Λ is $N(\theta,1)$, $-\infty < \theta < \infty$ is carried out and its power is compared with the classical UMP test and with $\delta_{\theta=o}$. For the case when Λ is binomial the powers of the classical UMP and the test $\delta_{\theta=o}$ for the size $\alpha = .05$ are compared.

To conclude it must be mentioned that a much more concise version of Meeden's work appeared in the Annals of Mathematical Statistics, June 1970, when the present thesis was ready for typing.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Professor S.G. Ghurye for introducing me to the problem and guiding me throughout the preparation of this thesis. I would also like to thank Professor E.S. Keeping for reading the manuscript and making many helpful comments, which led to a considerable improvement of the final version of the thesis. Thanks are also due to Dr. B.J. Mailloux for reading the manuscript. Mrs. Mary Willard, Programmer Analyst, helped to prepare some tables and Mrs. Vivian Spak did the necessary typing.

Finally I wish to express my gratitude to the University of Alberta and the University of Ife for their financial assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND RECENT DEVELOPMENTS

1.1 An Outline of Developments in Bayesian Testing

The problem of Bayesian testing of hypotheses is one where the parameter under test is assumed to be a random variable. The classical theory assumes the parameter to be an unknown constant. For some time one school ignored the possibility of the parameter being a random quantity. Advances were made in the classical field, but for the Bayesian school there were very few significant developments until 1950.

During the period 1915-1920 S. Bernstein and R. Von Mises independently proved that if the a priori density of a random variable θ is continuous, then as the number of independent observations is increased the a posteriori distribution of θ tends to a limit distribution which is independent of the a priori density. However, in some applications the number of observations is severely limited. Recently (1950-) some developments have taken place in testing when the parameter is a random variable.

In 1950 H. Robbins considered what he termed the compound decision problem. Essentially he considered the testing of a simple hypothesis against a simple alternative; $\Lambda = -1$ or $+1$ and $f(x|\lambda) = N(\lambda,1)$. Considering both errors equally serious, he exhibited a decision rule which, compared to the classical MP tests, reduced the overall error in the limit (as sample size $\to \infty$). This rule has

some unfortunate implications. In the following section, his paper of 1950 is briefly described.

1.2 The Compound Decision Problem

1.2.1 Statement of the Compound Decision Problem

Let $\underline{x} = (x_1,\ldots,x_n)$ be a random vector with x_i $(i=1,\ldots,n)$ independent normal random variables with common variance 1 and respective means $\theta_1,\ldots,\theta_n$ where $\theta_i = +1$ or -1, $i = 1,\ldots,n$. On the basis of the observed $\underline{x}$ we are to decide whether $\theta_i = 1$ or -1, $i = 1,\ldots,n$, in such a way as to minimize the expected total number of errors.

Let $\Omega = \{\underline{\theta}:\underline{\theta}=(\theta_1,\ldots,\theta_n),\ \theta_i = -1 \text{ or } 1,\ i=1,\ldots,n\}$. Then for a $\underline{\theta} \in \Omega$ $\underline{x}$ has density:

$$\phi(\underline{x},\underline{\theta}) = \frac{1}{(2\pi)^{n/2}} \exp[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_i)^2]$$

$$= \frac{1}{(2\pi)^{n/2}} \exp[-\frac{1}{2}(\chi^2+n)]\ \exp[\underline{\theta}\cdot\underline{x}^T] \qquad (1)$$

1.2.2 Loss Function and Decision Rule R

Let $w(\theta',\underline{\theta}) = \frac{1}{n}(\#i \ni \theta_i' \neq \theta_i)$

$$= \frac{1}{2n}\sum_{i=1}^{n}|\theta_i'-\theta_i| \qquad (2)$$

Let $\underline{\theta}^{(1)},\ldots,\underline{\theta}^{(2^n)}$ be an arbitrary ordering of Ω. Then

R, a decision rule (randomised) specifies a probability distribution, given $\underline{x}$, i.e. the probability that $\underline{\theta} = \underline{\theta}^{(j)}$, given $\underline{x} = p(\underline{\theta}^{(j)}|\underline{x})$; for convenience denote this probability by $p_j(\underline{x})$, then

$$R : p_j(\underline{x}) \quad j = 1,\ldots,2^n \quad p_j(x) \geq 0 \text{ and } \sum_j p_j(x) = 1 \tag{3}$$

If $L(R,\underline{\theta})$ is the expected loss in using R where the true parameter vector is $\underline{\theta}$ then $L(R,\underline{\theta}) = p(\underline{\theta}) - \frac{1}{n}\sum_{i=1}^{n} \text{Sgn}\,(\theta_i) \int \phi(\underline{x},\underline{\theta})U_i(\underline{x})d\underline{x}$ where

$$p(\theta) = \frac{1}{2n}\sum_{i=1}^{n}(1+\theta_i) \quad 0 \leq p(\underline{\theta}) \leq 1 \tag{4}$$

and

$$U_i(\underline{x}) = \frac{1}{2}\sum_j p_j(\underline{x})(1+\theta_i^{(j)}) \tag{5}$$

For, expected loss given $\underline{\theta}$ and R for any $\underline{x}$,

$$= \frac{1}{2n}\sum_{j=1}^{2^n}\sum_{i=1}^{n}|\theta_i^{(j)}-\theta_i|\int p(\underline{\theta}^{(j)}|\underline{x})\phi(\underline{x},\underline{\theta})d\underline{x}$$

$$= \frac{1}{2n}\sum_{i=1}^{n}\int[\sum_{j=1}^{2^n} p_j(\underline{x})|\theta_i^{(j)}-\theta_i|]\phi(\underline{x},\underline{\theta})d\underline{x}$$

$$\text{Now} \quad \sum_j p_j(\underline{x})\,|\theta_i^{(j)} - \theta_i| = \begin{cases} 2U_i(x) & \text{if} \quad \theta_i = -1 \\ 2[1-U_i(\underline{x})] & \text{if} \quad \theta_i = +1 \end{cases}$$

$$= 1 + \theta_i - 2\,\mathrm{Sgn}(\theta_i)U_i(\underline{x}) \quad \text{for} \quad \theta_i = \mp 1$$

Hence

$$L(R,\underline{\theta}) = p(\underline{\theta}) - \frac{1}{n}\sum_{i=1}^{n} \mathrm{Sgn}(\theta_i) \int \phi(\underline{x},\underline{\theta})U_i(\underline{x})d\underline{x} \tag{6}$$

Thus, in general, $L(R,\underline{\theta})$ can be expressed in terms of $U_i(\underline{x})$ $i = 1,\ldots,n$.

1.2.3 Rule $\tilde{R}$

Let

$$\tilde{R} : \theta_i = \mathrm{Sgn}(x_i) \quad i = 1,\ldots,n . \tag{7}$$

Then $L(\tilde{R},\underline{\theta}) = 0.1587$ $\forall\underline{\theta} \in \Omega$. For $U_i(\underline{x})$ is $p\{\theta_i=1|\underline{x}\}$. Indeed

$$\begin{aligned} p\{\theta_i = 1|\underline{x}\} &= \sum_{\{\underline{\theta}:\theta_i=1\}} p_j(\underline{x}) \\ &= \sum_{\Omega} p_j(\underline{x}) \frac{(1+\theta_i^{(j)})}{2} \\ &= U_i(\underline{x}). \end{aligned}$$

$U_i(\underline{x})$ is also the expected value, given $\underline{x}$, of $\frac{1+\theta_i^{(j)}}{2}$ under $\tilde{R}$. Hence $U_i(\underline{x})$ is the conditional probability, given $\underline{x}$, of $\theta_i = 1$. So for $\tilde{R}$

$$U_i(x) = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{if } x_i < 0 \end{cases}$$

$$\therefore \quad L(\tilde{R},\underline{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1+\theta_i}{2} - \mathrm{Sgn}(\theta_i) \int_{x_i>0} \phi d\underline{x}\right]$$

which after integrating with respect to variables other than x_i, and then substituting $x_i - \theta_i = y$ becomes

$$L(\tilde{R},\underline{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1+\theta_i}{2} - \frac{\mathrm{Sgn}(\theta_i)}{\sqrt{2\pi}} \int_{-\theta_i}^{+\infty} \exp[-\tfrac{1}{2}y^2]dy\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1+\theta_i}{2} - \mathrm{Sgn}(\theta_i)[1-F(-\theta_i)]\right]$$

$$= F(-1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1} \exp[-\tfrac{1}{2}y^2]dy$$

$$= .1587 \quad \forall \underline{\theta} \in \Omega .$$

1.2.4 A Theorem and Decision Rule $\bar{R}$

Theorem 1.2.4.1. Let

(i) R be a decision rule with the associated functions $U_i(\underline{x})$ $i = 1,\ldots,n$ as in (5) .

(ii) $\Omega_k = \{\underline{\theta}:\underline{\theta}\varepsilon\Omega,\ p(\underline{\theta}) = \frac{k}{n}\}$ $k = 0,1,\ldots,n$. Thus $\underline{\theta} \in \Omega_k$ iff exactly k components of $\underline{\theta}$ are 1 .

(iii) $h(\underline{\theta}) \geq 0,\ \not\equiv 0$ be defined on Ω such that $h(\underline{\theta})$ = constant = b_k on Ω_k $k = 0,\ldots,n$.

Then for a given $h(\underline{\theta})$, $\sum_{\Omega} h(\underline{\theta})L(R,\underline{\theta})$ = min for the (non-randomised) rule

$$R : \theta_i = \mathrm{Sgn}\left(x_i - \frac{1}{2} \ln \frac{\sum_{k=o}^{n-1} b_k S_k^{(i)}}{\sum_{k=o}^{n-1} b_{k+1} S_k^{(i)}} \right)$$

where $S_k^{(i)} = \sum e^{2(x_{j_1} + x_{j_2} + \ldots + x_{j_k})}$ $k = 1,\ldots,(n-1)$ $S_{-1}^{(i)} = S_n^{(i)} = 0$, $S_o^{(i)} = 1$ $\forall i$ and $\sum$ is taken over all $\binom{n-1}{k}$ combinations of the integers $1,2,\ldots,i-1,i+1,\ldots,n$. k at a time.

Proof: As seen in (6), $L(R,\underline{\theta}) = p(\underline{\theta}) - \frac{1}{n} \sum_{i=1}^{n} \mathrm{Sgn}(\theta_i) \int \phi(\underline{x},\underline{\theta}) U_i(\underline{x}) d\underline{x}$

$$\therefore \sum_{\Omega} h(\underline{\theta}) L(R,\underline{\theta}) = \sum_{\Omega} h(\underline{\theta}) p(\underline{\theta}) - \frac{1}{n} \sum_{i=1}^{n} \int [\sum_{\Omega} h(\underline{\theta})\ \mathrm{Sgn}(\theta_i)\phi] U_i(\underline{x}) d\underline{x}$$

This will be min iff

$$U_i(\underline{x}) = \begin{cases} 1 & \text{if } \sum_{\Omega} h(\underline{\theta})\ \mathrm{Sgn}(\theta_i)\phi(\underline{x},\underline{\theta}) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Let

$$\begin{aligned} \Omega_{k,i}^{+} &= \{\underline{\theta} : \underline{\theta} \varepsilon \Omega_k,\ \ \theta_i = 1\} \\ \Omega_{k,i}^{-} &= \{\underline{\theta} : \underline{\theta} \varepsilon \Omega_k,\ \ \theta_i = -1\} \end{aligned} \tag{10}$$

Thus $\Omega_k = \Omega_{k,i}^{+} + \Omega_{k,i}^{-}$, where $+$: Union. Then, from (9), $U_i(x) = 1$ if

$$\sum_{k=o}^{n} b_k [\sum_{\Omega^+_{k,i}} \phi(\underline{x},\underline{\theta}) - \sum_{\Omega^-_{k,i}} \phi(\underline{x},\underline{\theta})] > 0$$

But $\phi(\underline{x},\underline{\theta}) = \frac{1}{(2\pi)^{n/2}} \exp[-\frac{1}{2}(\chi^2+n)] \exp[\underline{\theta}\cdot\underline{x}^T]$, see (1) Hence

$U_i(\underline{x}) = 1$ if

$$\sum_{k=o}^{n} b_k [\sum_{\Omega^+_{k,i}} e^{\underline{\theta}\cdot\underline{x}^T} - \sum_{\Omega^-_{k,i}} e^{\underline{\theta}\cdot\underline{x}^T}] > 0$$

or multiplying throughout by $e^{\underline{1}\cdot\underline{x}^T}$, where $\underline{1} = (1,1,\ldots,1)$, we have:

$$\sum_{k=o}^{n} b_k [\sum_{\Omega^+_{k,i}} e^{(\underline{1}+\underline{\theta})\cdot\underline{x}^T} - \sum_{\Omega^-_{k,i}} e^{(\underline{1}+\underline{\theta})\cdot x^T}] > 0$$

Now for $\Omega^+_{k,i}$, $\theta_i = 1$ and hence

$$\sum_{\Omega^+_{k,i}} e^{(\underline{1}+\underline{\theta})\cdot\underline{x}^T} = \sum_{\Omega^+_{k,i}} e^{2x_i} \cdot e^{2(x_{j_1}+x_{j_2}+\ldots+x_{j_{k-1}})}$$

where j_r , $r = 1,\ldots,(k-1)$, belong to $\{1,2,\ldots,i-1,i+1,\ldots,n\}$.

Hence

$$\sum_{\Omega^+_{k,i}} e^{(\underline{1}+\underline{\theta})\cdot\underline{x}^T} = e^{2x_i} S^{(i)}_{k-1} .$$

Similarly for $\Omega^{-}_{k,i}$, $\theta_i = -1$ and hence

$$\sum_{\Omega^{-}_{k,i}} e^{(\underline{1}+\underline{\theta})\cdot\underline{x}^T} = \sum_{\Omega^{-}_{k,i}} e^{2(x_{j_1}+x_{j_2}+\ldots+x_{j_k})}$$

$$= S_k^{(i)}$$

or $U_i(\underline{x}) = 1$ if

$$\sum_{k=o}^{n} b_k[e^{2x_i}S_{k-1}^{(i)} - S_k^{(i)}] > 0$$

or

$$x_i > \frac{1}{2} \ln \frac{\sum\limits_{k=o}^{n-1} b_k S_k^{(i)}}{\sum\limits_{k=o}^{n-1} b_{k+1} S_k^{(i)}} \tag{11}$$

Hence $\sum\limits_{\Omega} h(\underline{\theta})L(R,\underline{\theta}) = \min$ for the rule

$$R : \theta_i = \mathrm{Sgn}\left(x_i - \frac{1}{2} \ln \frac{\sum\limits_{o}^{n-1} b_k S_k^{(i)}}{\sum\limits_{o}^{n-1} b_{k+1} S_k^{(i)}}\right) \tag{12}$$

Definition: A decision rule R is called symmetric if $L(R,\underline{\theta}) =$ constant $= c_k$ for $\underline{\theta} \in \Omega_k$ $k = 0,1,\ldots,n$.

Corollary: 1.2.4.1. If $b_k = 1 \ \forall k = 0,1,\ldots,n$.

$$R = \tilde{R} : \ \theta_i = \mathrm{Sgn}(x_i)$$

and $\sum_{\Omega} L(R,\underline{\theta}) = \min$ when $R = \tilde{R}$. Then for any R with associated functions $U_i(\underline{x})$

$$\sum_{\Omega} L(R,\underline{\theta}) > \sum_{\Omega} L(\tilde{R},\underline{\theta}) = 2^n F(-1)$$

Hence $\max_{\underline{\theta}\varepsilon\Omega} L(R,\underline{\theta}) > \max_{\underline{\theta}\varepsilon\Omega} L(\tilde{R},\underline{\theta}) = F(-1)$.

Corollary: 1.2.4.2. If $b_k = \binom{n}{k}^{-1}$ then by (12) we have that $\sum_{k=o}^{n} \binom{n}{k}^{-1} \sum_{\Omega_k} L(R,\underline{\theta}) = \min$ in the class of all R , for the symmetric rule

$$\bar{R} : \theta_i = \mathrm{Sgn}\left(x_i - \frac{1}{2} \ln \frac{\sum_{k=o}^{n-1} \binom{n}{k}^{-1} S_k^{(i)}}{\sum_{k=o}^{n-1} \binom{n}{k+1}^{-1} S_k^{(i)}}\right) \tag{13}$$

To prove that $\bar{R} : \theta_i = \mathrm{Sgn}\left(x_i - \frac{1}{2} \ln \frac{\sum_{k=o}^{n-1} \binom{n}{k}^{-1} S_k^{(i)}}{\sum_{k=o}^{n-1} \binom{n}{k+1}^{-1} S_k^{(i)}}\right)$ is symmetric, we prove that $\forall\underline{\theta} \ \varepsilon \ \Omega_k$, $k = 0,1,2,\ldots,n$ $L(\bar{R},\underline{\theta}) = c_k$.

Proof: Let $\underline{\theta} \ \varepsilon \ \Omega_k$ and let

$$\frac{1}{2} \ln \frac{\sum_{k=o}^{n-1} \binom{n}{k}^{-1} S_k^{(i)}}{\sum_{k=o}^{n-1} \binom{n}{k+1}^{-1} S_k^{(i)}} = L(i)$$

$$L(\overline{R},\underline{\theta}) = \frac{1}{2n} \sum_{j=1}^{2^n} \sum_{i=1}^{n} |\theta_i^{(j)}-\theta_i| \int p(\underline{\theta}^{(j)}|\underline{x})\phi(\underline{x},\underline{\theta})d\underline{x}$$

$$= \frac{1}{2n} \sum_{s=o}^{n} \sum_{\Omega_s} \sum_{i=1}^{n} |\theta_i^{(j)}-\theta_i| \int p(\underline{\theta}^{(j)}|\underline{x})\phi d\underline{x} \tag{13a}$$

Suppose $\underline{\theta}^{(j)} \varepsilon \Omega_s$ where $s \varepsilon \{0,1,2,...,n\}$; suppose $\theta_{j_r}^{(j)} = 1$ for $r = 1,2,...,s$ and $\theta_{j_r}^{(j)} = -1$ for $r = s + 1,...,n$. Then given $\underline{x}$

$$p(\underline{\theta}^{(j)}|\underline{x}) = 1 \quad \text{if} \quad x_{j_r} - L(j_r) > 0 \;, \quad r = 1,...,s$$

$$\text{and} \quad x_{j_r} - L(j_r) < 0 \;, \quad r = s+1,...,n$$

$$= 0 \quad \text{otherwise.}$$

Thus $\int p(\underline{\theta}^{(j)}|\underline{x})\phi d\underline{x} = \int_S \phi d\underline{x}$

where $S = \{\underline{x}: x_{j_r} - L(j_r) > 0 \quad r = 1,2,...,s$ and

$x_{j_r} - L(j_r) < 0 \quad r = s+1,...,n\}$.

This is the same $\forall\underline{\theta} \varepsilon \Omega_k$ and $\forall\underline{\theta} \varepsilon \Omega_s$. Let this be $p(s,k)$; then by (13a)

$$L(\bar{R},\underline{\theta}) = \frac{1}{2n}\sum_{s=o}^{n}\left[\sum_{\underline{\theta}\varepsilon\Omega_s}\sum_{i=1}^{n}|\theta_i^{(j)}-\theta_i|p(s,k)\right]$$

$$= \frac{1}{2n}\sum_{s=o}^{n}p(s,k)\sum_{i=1}^{n}\left[\sum_{\Omega_s}|\theta_i^{(j)}-\theta_i|\right]$$

$$= \frac{1}{2n}\sum_{s=o}^{n}p(s,k)\sum_{i=1}^{n}\left[\binom{n-1}{s}|-1-\theta_i|+\binom{n-1}{s-1}|1-\theta_i|\right]$$

$$= \frac{1}{2n}\sum_{s=o}^{n}p(s,k)\left[k\binom{n-1}{s}\cdot 2+(n-k)\binom{n-1}{s-1}\cdot 2\right]$$

which depends only on k and hence $\forall\underline{\theta}\ \varepsilon\ \Omega_k$, $L(\bar{R},\underline{\theta}) = c_k$, $k = 0,1,\ldots,n$. □

Also note that if R is any symmetric rule with $L(R,\underline{\theta}) = c_k$, and if $\underline{\theta}\ \varepsilon\ \Omega_k$ then $\sum_{k=o}^{n} c_k = \sum_{k=o}^{n}\binom{n}{k}^{-1}\sum_{\Omega_k} L(R,\underline{\theta})$ as there are $\binom{n}{k}$ points in Ω_k. As a result if $L(\bar{R},\underline{\theta}) = \bar{c}_k$, $\underline{\theta}\ \varepsilon\ \Omega_k$ then

$$\sum_{k=o}^{n}\bar{c}_k < \sum_{k=o}^{n}\tilde{c}_k = (n+1)F(-1)$$

and

$$\frac{1}{n+1}\sum_{k=o}\bar{c}_k < F(-1) .$$

1.2.5 The Rule R^*

Let

$$x^* = \begin{cases} +\infty & \text{if } \bar{x} \leq -1 \\ \frac{1}{2}\ln\frac{1-\bar{x}}{1+\bar{x}} & \text{if } -1 < \bar{x} < 1 \\ -\infty & \text{if } \bar{x} \geq 1 \end{cases} \qquad (14)$$

and let

$$R^* : \theta_i = \mathrm{Sgn}(x_i - x^*) \quad i = 1,\dots,n \ . \tag{15}$$

Then

$$L(R^*,\underline{\theta}) = h[p(\underline{\theta}),n] \tag{16}$$

and

$$\lim_{n\to\infty} \{L(R^*,\underline{\theta}) - h(p(\underline{\theta}))] = 0 \quad \text{uniformly} \quad \forall\underline{\theta} \in \Omega \tag{17}$$

where

$$\begin{aligned} h(p,n) &= pF(-2p\sqrt{n}) + (1-p)F[-2(1-p)\sqrt{n}] \\ &+ \int_{-2(1-p)\sqrt{n}}^{2p\sqrt{n}} \{pF[\sqrt{\frac{n}{n-1}}\ (-1-\frac{x}{\sqrt{n}} + \frac{1}{2}L(p))] \\ &+ (1-p)F[\sqrt{\frac{n}{n-1}}\ (-1-\frac{x}{\sqrt{n}} - \frac{1}{2}L(p))]\}f(x)dx \end{aligned} \tag{18}$$

where $f(x) = N(0,1)$, $L(p) = \frac{1}{2}\ln\dfrac{1-p+\frac{x}{2\sqrt{n}}}{p-\frac{x}{2\sqrt{n}}}$ and

$$h(p) = pF(-1+\frac{1}{2}\ln\frac{(1-p)}{p}) + (1-p)F(-1-\frac{1}{2}\ln\frac{1-p}{p}) \ . \tag{19}$$

Robbins compares the risks for R^* and $\tilde{R}$ for $n = 100$. The table is as follows:

TABLE I

p	$L(\tilde{R},\underline{\theta})$	$L(R^*,\underline{\theta})$	h(p)
0 or 1	.1587	.0041	0
.1 or .9	.1587	.0763	.0691
.2 or .8	.1587	.1174	.1121
.3 or .7	.1587	.1439	.1387
.4 or .6	.1587	.1591	.1538
.5	.1587	.1628	.1587

Here
$$p = \frac{1}{n} (\# \text{ of } \theta_i \ni \theta_i = 1)$$

Hence if $p \simeq .5$ then, considering the expected number of errors, it is better to use $\tilde{R}$; however if $p \not\simeq .5$, then by using R^* one would be balancing the possibility of a slightly higher risk near $p \simeq .5$ against a much lower risk near $p = 0$ or 1 .

In conclusion R^* is not shown here to be a best rule in any sense, but it is as simple as $\tilde{R}$ and in certain circumstances $(p \not\simeq .5)$ it would seem better than $\tilde{R}$.

1.2.6 Implications of Procedure R^*

The use of R^* to test $\theta_i = +1$ or -1 $i = 1,\ldots,n$ implies that the decision, say, to accept or reject each of n products (one of which may be a pair of shoes, another a shirt and yet another the height of an individual) depends on all the products. And this

aspect is unsatisfactory say to the manufacturer of shoes. However it is argued that if R^* reduces the cost of sampling, and this reduction is passed on to the consumer it may be welcome to him.

1.2.7 Asymtotic Solutions of the Compound Decision Problem

In 1955 J. Hannan and H. Robbins considered the same problem when $f(x|\lambda)$ was any completely specified distribution, $\lambda = 0$ or 1. They took the risk to be the weighted probability of wrong decisions, and, for n sufficiently large, they found decision functions whose risks are uniformly close to the envelope risk function of "invariant" decision functions.

In 1963 E. Samuel proved, for testing a composite hypothesis against a composite alternative, the convergence of the risk function of empirical Bayes rules to the Bayes envelope risk function for geometric, Poisson and negative binomial distributions. Then Samuel also considered tests where x was a continuous random variable with $f(x|\lambda)$ of the exponential family. The results on discrete distributions are summarized in the following section.

1.3 <u>The Empirical Bayes Approach</u>

1.3.1 Statement of the Problem and Notation

We consider a situation where $\{(X_i,\Lambda_i)\}_{i=1}^{n}$ is a sequence of independent pairs of random variables such that the Λ_i's are identically distributed with common distribution function G and where for $i = 1,\dots,n$ $P\{X_i \leq x|\lambda\} = P[x|\lambda]$. Further the pairs (X_i,Λ_i)

arise sequentially and after each occurrence the statistician has to take a decision about Λ, which is not observable.

We are concerned with testing

$$H_o : \lambda \le \lambda_o \quad \text{against} \quad H_1 : \lambda > \lambda_o \tag{19}$$

$$H_o : \lambda \ge \lambda_o \quad \text{against} \quad H_1 : \lambda < \lambda_o \tag{20}$$

$$H_o : |\lambda-\lambda_o| \le \Delta \quad \text{against} \quad H_1 : |\lambda-\lambda_o| > \Delta \tag{21}$$

$$H_o : |\lambda-\lambda_o| \ge \Delta \quad \text{against} \quad H_1 : |\lambda-\lambda_o| < \Delta \tag{22}$$

Furthermore we deal with the Poisson, geometric and negative binomial distributions where the probability density, given λ, is as follows:

$$P(x|\lambda) = \lambda^x h(\lambda) g(x) \qquad x = 0,1,2,\ldots \tag{23}$$

where for:

Poisson $\quad h(\lambda) = e^{-\lambda}, \quad g(x) = \frac{1}{x!} \quad \lambda > 0$

geometric $\quad h(\lambda) = \frac{1}{(1-\lambda)}, \quad g(x) = 1 \quad (0<\lambda<1)$

negative binomial $\quad h(\lambda) = (1-\lambda)^{-m}, \quad g(x) = \binom{x+m-1}{x}$, m a positive integer $0 < \lambda < 1$.

Since in (19)-(22) we have two-action problems let

$$A_i : H_i \text{ is correct} \quad i = 0,1 \, . \tag{24}$$

and the loss incurred in taking action A_i, when λ is the true parameter value, be

$$L_i = L[A_i,\lambda] \quad i = 0,1 \, . \tag{25}$$

Then a decision function $t(\underline{x})$ is such that when $\underline{X} = \underline{x}$ is observed we decide with probability $t(\underline{x})$ that A_1 is true and with probability $1 - t(\underline{x})$ that A_o is true.

Definition:

$$R(t,G) = \int_\Omega \int_{\mathcal{X}} \{t(x)L_1+(1-t(x))L_o\} f d\mu dG$$

$$= E_G[L_o] - E\{t(x)E_G[L_o-L_1|x]\} \tag{26}$$

where $f(x|\lambda)$ is the conditional density of x given λ with respect to some measure μ . E denotes expectation with respect to the distribution of (X,Λ), E_G to that of Λ and $E_G[\ |x]$ conditional expectation given x.

To prove (26) we have

$$R(t,G) = \int_\Omega \int_{\mathcal{X}} \{t(x)L_1+(1-t(x))L_o\} f(x|\lambda) d\mu dG$$

$$= \int_\Omega \int_{\mathcal{X}} L_o f(x|\lambda) d\mu dG - \int_\Omega \int_{\mathcal{X}} t(x)[L_o-L_1] f(x|\lambda) d\mu dG$$

$$= E_G[L_o] - \int_{\mathcal{X}} t(x)[\int_\Omega [L_o-L_1|x] f(\lambda|x) d\nu dG$$

where $f(\lambda|x)$ is the conditional density with respect to some measure ν .

$$R(t,G) = E_G[L_o] - \int_{\mathcal{X}} t(x)E_G[L_o-L_1|x]f(x)dx$$

$$= E_G[L_o] - E\{E_G[L_o-L_1|x]\}. \square$$

For a given G, there usually exists a decision rule ϕ_G such that

$$R(\phi_G) = \min_{\phi} R(\phi,G) .$$

From (26) for a rule ϕ

$$R(\phi,G) = E_G[L_o] - E\{t(x)E_G[L_o-L_1|x]\}$$

and ϕ's which minimize $R(\phi,G)$ over rules ϕ are of the form

$$\begin{aligned} t_G(x) &= 1 && \text{if } E_G[L_o-L_1|x] > 0 \\ &= 0 && \text{if } E_G[L_o-L_1|x] < 0 \\ &= \text{arb}[0,1] && \text{if } E_G[L_o-L_1|x] = 0 . \end{aligned} \tag{27}$$

Definition: The $\min_{\phi} R(\phi,G)$ is called the Bayes' envelope function. Denote this by $R(G)$.

Let $\underline{X}_n$ be a random vector independent of vector (X,Λ), $k_n(x,\underline{X}_n)$ a function of $(x,\underline{X}_n)$ and $k(x)$ and of x only. Define

$$t_n^k(x,\underline{X}_n) = 1 \quad \text{if} \quad k_n(x;\underline{X}_n) > 0$$
$$= 0 \quad \text{if} \quad k_n(x;\underline{X}_n) \leq 0 \tag{28}$$

$$t^k(x) = 1 \quad \text{if} \quad k(x) > 0$$
$$= 0 \quad \text{if} \quad k(x) \leq 0 \tag{29}$$

$$\begin{aligned} t_*^k(x) &= 1 && \text{if} \quad k(x) > 0 \\ &= 0 && \text{if} \quad k(x) < 0 \\ &= \text{arbitrary } [0,1] && \text{if} \quad k(x) = 0 \ . \end{aligned} \tag{30}$$

Suppose $k(x)$ is such that

$$R(t^k,G) = R(t_*^k,G) \quad \forall t_*^k \tag{31}$$

This is satisfied if $k(x) = E_G[L_o - L_1 | x]$.

1.3.3 A Theorem

Theorem: If (i) $k_n(x,\underline{X}_n) \to k(x)$ in probability

(ii) $E_G[|L_o - L_1|] < \infty$

Then
$$\begin{aligned} \lim_{n\to\infty} R(t_n^k,G) &= R(t^k,G) \\ &= R(G) \quad \text{if} \quad k(n) = E_G[L_o - L_1 | x] \\ &= \text{Bayes envelope risk.} \end{aligned}$$

Proof: We use:

E to denote expectation with respect to the distribution of $(X,\Lambda),\underline{X}_n$

E_* to denote expectation with respect to the distribution of (X,Λ)

E_n to denote expectation with respect to the distribution of $\underline{X}_n$

E_G to denote expectation with respect to the distribution of Λ

$E_G[\,|x]$ the conditional expectation given X .

We have to show that:

$$\lim_{n\to\infty} E\{t_n^k(X,\underline{X}_n)\cdot E_G[L_o-L_1|x]\} = E_*\{t^k(X)E_G[L_o-L_1|x]\}$$

Now $E\{t_n^k(X,\underline{X}_n)E_G[L_o-L_1|x]\} = E_*\{E_n[t_n^k(X,\underline{X}_n)|x]\cdot E_G[L_o-L_1|x]$

$$\lim_{n\to\infty} E_n\{t_n^k(x,\underline{X}_n)\} \to t^k(x) \quad \text{by definition of} \quad t_n^k(x,\underline{X}_n) \quad \text{in} \quad (28)$$

and be hypothesis (i).

$$\therefore \lim_{n\to\infty} E_n[t_n^k(x,\underline{X}_n)]E_G[L_o-L_1|x] \to t^k(x)E_G[L_o-L_1|x] .$$

Also $|E_n[t_n^k(X,\underline{X}_n)]E_G[L_o-L_1|x]| \leq E_G[|L_o-L_1||x]$ as $|t_n^k| \leq 1$. Hence using hypothesis (ii) and the Lebesgue dominated convergence theorem

$$\lim_{n\to\infty} E_*\{E_n[t_n^k(X,\underline{X}_n)E_G[L_o-L_1|x]\} = E_*\{t^k(X)E_G[L_o-L_1|x]\}$$

i.e. $\lim_{n\to\infty} E\{t_n^k(X,\underline{X}_n)E_G[L_o-L_1|x]\} = E_*\{t^k(X)E_G[L_o-L_1|x]\}$. $\square$

1.3.4 Application to the Poisson, Geometric and Negative Binomial Distributions

To test (19) let

$$L_o = c(\lambda-\lambda_o) \quad \lambda > \lambda_o$$
$$= 0 \quad \quad \leq \lambda_o$$

$$L_1 = c(\lambda_o - \lambda) \qquad \lambda < \lambda_o$$
$$= 0 \qquad \geq \lambda_o$$

where $c > 0$. Then in case of the geometric and negative binomial $0 \leq \lambda \leq 1$ and, whatever G, we have $E_G[|L_o - L_1|] < \infty$. For the Poisson case, we assume that $E_G[|\lambda|] < \infty$. Hence, for all three distributions, (ii) is satisfied.

Now define

$$P_n(x) = P_n(x, \underline{X}_n) \qquad x = 0,1,2,\ldots$$
$$= \frac{1}{n}\{\# \text{ of indices } i \text{ such that } X_i = x \quad i = 1,\ldots,n\} .$$

Then, by the strong law of large numbers,

$$\lim_{n\to\infty} P_n(x) = P_G(x) \tag{32}$$

the marginal density of X whatever G. Now

$$E_G[L_o - L_1 | x] = \frac{c\int_\Omega (\lambda - \lambda_o) P(x|\lambda) dG(\lambda)}{\int_\Omega P(x|\lambda) dG(\lambda)} \tag{33}$$

By (23) $P_G(x) = g(x) \int_\Omega \lambda^x h(\lambda) dG(\lambda)$. Hence

$$E_G[L_o - L_1 | x] = c\,\frac{g(x)}{g(x+1)}\,\frac{P_G(x+1)}{P_G(x)} - c\lambda_o \tag{34}$$

Letting $k_n(x,\underline{X}_n) = c\dfrac{g(x)}{g(x+1)}\dfrac{P_n(x+1)}{P_n(x)} - c\lambda_o$ by (32)

$k_n(x,\underline{X}_n) \to c\ \dfrac{g(x)}{g(x+1)}\ \dfrac{P_G(x+1)}{P_G(x)}$ in probability. Hence if we take $x = x_{n+1}$, the theorem applies and if

$$t_{n+1}(\underline{x}_{n+1}) = 1 \quad k_n(\underline{x}_{n+1}) > 0$$
$$= 0 \quad \text{otherwise.}$$

Then $\lim\limits_{n\to\infty} R(t_n,G) = R(G)$. Now

$$\begin{aligned} g(x) &= \frac{1}{x!} && \text{for Poisson} \\ &= 1 && \text{for geometric} \\ &= \binom{x+m-1}{x} \quad m > 0 && \text{negative binomial.} \end{aligned} \tag{35}$$

Hence the rules are $t_{n+1}(\underline{x}_{n+1}) = 1$ if

$$\begin{aligned} (x_{n+1}+1)P_n(x_{n+1}+1) &> \lambda_o P_n(x_{n+1}) && \text{Poisson} \\ P_n(x_{n+1}+1) &> \lambda_o P_n(x_{n+1}) && \text{geometric} \\ (x_{n+1}+1)P_n(x_{n+1}+1) &> \lambda_o(x_{n+1}+m)P_n(x_{n+1}) && \text{negative binomial} \end{aligned} \tag{36}$$

Similarly, rules for testing other hypotheses, (20)-(22), can be written down. For (21) i.e. $H_o : |\lambda-\lambda_o| \le \Delta$ against $H_1 : |\lambda-\lambda_o| > \Delta$. Let

$$\begin{aligned} L(A_o,\lambda) &= 0 && \text{if } |\lambda-\lambda_o| \le \Delta \\ &= c[(\lambda-\lambda_o)^2-\Delta^2] && \text{if } |\lambda-\lambda_o| > \Delta \\ L(A_1,\lambda) &= 0 && \text{if } |\lambda-\lambda_o| > \Delta \\ &= c[\Delta^2-(\lambda-\lambda_o)^2] && \text{if } |\lambda-\lambda_o| \le \Delta \end{aligned} \tag{37}$$

For the Poisson case we assume $E_G[\lambda^2] < \infty$. Then

$$E_G[L_o - L_1 | x] = \frac{cg(x) \int_\Omega [\lambda^2 - 2\lambda_o\lambda + (\lambda_o^2 - \Delta^2)]\lambda^x h(\lambda)\,dG(\lambda)}{g(x) \int_\Omega h(\lambda)\lambda^x dG(\lambda)}$$

$$= c\left\{\frac{g(x)}{g(x+2)} \frac{P_G(x+2)}{P_G(x)} - 2\lambda_o \frac{g(x)}{g(x+1)} \frac{P_G(x+1)}{P_G(x)}\right\}$$

$$+ c(\lambda_o^2 - \Delta^2) \tag{38}$$

using (23). Hence by (32) $k_n(x,\underline{x}_n) \to E_G[L_o - L_1 | x]$ where

$$k_n(x,\underline{x}_n) = c\left\{\frac{g(x)}{g(x+2)} \frac{P_n(x+2)}{P_n(x)} - 2\lambda_o \frac{g(x)}{g(x+1)} \frac{P_n(x+1)}{P_n(x)}\right\} + c(\lambda_o^2 - \Delta^2) \tag{39}$$

Finally taking $x = x_{n+1}$ we have $t_{n+1}(\underline{x}_{n+1}) = 1$ if, the following holds and 0 otherwise:

$$(x_{n+1}+2) \frac{P_n(x_{n+1}+2)}{P_n(x_{n+1})} - 2\lambda_o(x_{n+1}+1) \frac{P_n(x_{n+1}+1)}{P_n(x_{n+1})} > \Delta^2 - \lambda_o^2 \qquad \text{Poisson}$$

$$\frac{P_n(x_{n+1}+2)}{P_n(x_{n+1})} - 2\lambda_o \frac{P_n(x_{n+1}+1)}{P_n(x_{n+1})} > \Delta^2 - \lambda_o^2 \qquad \text{Geometric}$$

$$\frac{(x_{n+1}+2)_2}{(x_{n+1}+m+1)_2} \frac{P_n(x_{n+1}+2)}{P_n(x_{n+1})} - \frac{(x_{n+1}+1)}{(x_{n+1}+m)} \frac{P_n(x_{n+1}+1)}{P_n(x_{n+1})} > \Delta^2 - \lambda_o^2 \qquad \text{Negative Binomial.}$$

1.4 An Outline of Meeden's Results

There have been others who considered the problem of finding optimum Bayesian tests. One that concerns us is Glen Meeden's. Essentially, in his main theorems, he investigates the problem of existence of UMP Bayesian tests with a priori distribution partially known. The parameter θ of the a priori density is assumed unknown but in a set Θ .

He seeks a test which, subject to bounding the more serious error, minimizes the other error among a class of decision functions. Initially he establishes some relations between the Bayesian and classical tests. His main result proves the existence of a UMP Bayesian test when the conditional density $f(x|\lambda)$ has a monotone likelihood ratio in x and the hypothesis and the alternative are both composite. Examples are given to show that the class is reasonably interesting.

CHAPTER II

MEEDEN'S RESULTS AND COMMENTS

2.1 Notation and the Statement of the Problem

2.1.1 Meeden's work is part of his doctoral thesis. To summarize his results the following notation is used. Let $(\mathfrak{X},\mathcal{A})$ and $(\Omega,\mathcal{F})$ be two measurable spaces with $\mathfrak{X}$ and Ω as respective ranges for the random variables X and Λ, and $\mathcal{A}$ and $\mathcal{F}$ σ-algebras of measurable sets of $\mathfrak{X}$ and Ω respectively. Let ω be an $\mathcal{F}$-measurable subset of Ω and ω^c its complement. Let $f(x|\lambda)$ $\lambda \varepsilon \Omega$ be the conditional density of the random variable X given $\Lambda = \lambda$ with respect to some σ-finite measure μ . Sometimes we need the conditional density of X given ω then

$$f(x|\omega) = \frac{\int_\omega f(x|\lambda)dG}{G(\omega)} \qquad G(\omega) > 0$$

Let $\mathcal{G} = \{G{:}G$ is a c.d.f. for Λ and G satisfies $P\}$ where P is some property. G will also be used as a measure on Ω when required.

The problem considered is this: (X,Λ) is a random vector but only X is observable and it is required to decide, using the observation on X, whether $\Lambda \varepsilon \omega$ or $\Lambda \varepsilon \omega^c$. Let $\delta(X)$ denote a decision rule used. Then $\delta(X)$ is required to be optimum as described below in Section 2.1.2. For the moment $\delta : \mathfrak{X} \to [0,1]$ such that when $X = x$ is observed it is decided with probability $\delta(x)$ that $\Lambda \varepsilon \omega^c$.
Thus $\delta(X)$ is a test of

$$H : \Lambda \in \omega \quad \text{against} \quad A : \Lambda \in \omega^c \tag{1}$$

Sometimes we need to refer to $\delta(X)$ as a test in the classical sense and sometimes in the Bayesian sense. Meeden uses the word "predictor" when $\delta(X)$ is considered as a Bayesian test. But "predictor" does not seem appropriate as it gives the incorrect impression that X occurs first and then Λ. Finally let $0 \leq \alpha \leq 1$ and

$$D^{\alpha}_{\mathcal{G}} = \{\delta(X) : \int_{\mathfrak{X}} \int_{\omega} \delta f(x|\lambda)\, dG d\mu \leq \alpha \quad \forall G \in \mathcal{G}\}.$$

2.1.2 The Statement of the Problem

Given a $\mathcal{G}$ the problem is to find a δ_o, if any exists, such that

$$\delta_o \in D^{\alpha}_{\mathcal{G}} \qquad \text{(a)}$$

and δ_o maximizes (2)

$$\int_{\mathfrak{X}} \int_{\omega^c} \delta f(x|\lambda)\, dG d\mu \quad \forall G \in \mathcal{G} \text{ and } \forall \delta \in D^{\alpha}_{\mathcal{G}} \qquad \text{(b)}$$

In order to avoid repetition whenever (2) has a solution relative to $\mathcal{G}$ [as specified in that context] we shall denote it by δ_o and may refer to it as the UMP Bayesian test relative to $\mathcal{G}$.

2.2 Some Relations Between the Classical Tests and the Bayesian Tests

(A) Let G_o be a known distribution for Λ and $\mathcal{G} = \{G_o\}$. Then (2) has a solution, by the Neyman-Pearson lemma; for both

$$\int_{\omega} f(x|\lambda)dG_o \quad \text{and} \quad \int_{\omega^c} f(x|\lambda)dG_o$$

are non-negative, μ-integrable functions and the lemma is true for such functions. Hence there is no real claim for extension.

(B) Let $\mathcal{G} = \{G:G(\omega) = \gamma,\ \text{known},\ 0<\gamma<1\}$. Then $\delta_o \Longleftrightarrow \delta_o$ UMP classical level $\min(\frac{\alpha}{\gamma},1)$ test of (1).

(C) Let $\mathcal{G} = \{G:G(\omega)=\gamma$, unknown but $f(x|\omega)$ and $f(x|\omega^c)$ are known$\}$. Then

$$\delta_o \Longleftrightarrow \gamma \int_{\mathfrak{X}} \delta_o f(x|\omega)d\mu \leq \alpha \quad \forall G \in \mathcal{G}$$

and maximizes $\int_{\mathfrak{X}} \delta f(x|\omega^c)d\mu \quad \forall G \in \mathcal{G}$ and $\forall \delta \in D^{\alpha}_{\mathcal{G}}$

$\Longleftrightarrow \delta_o$ MP classical level α test of $H : X$ has density $f(x|\omega)$ against $A : X$ has density $f(x|\omega^c)$. Meeden only asserts sufficiency ($\Longleftarrow$) whereas the result is both necessary and sufficient.

(D) If $\mathcal{G} = \{G:G$ is any possible distribution for $\Lambda\}$ then $\delta_o \Longleftrightarrow \delta_o$ UMP classical L—α test of (1) . Note this is also true if $\mathcal{G} = \{G:G$ is a degenerate distribution for $\Lambda\}$.

2.3 Meeden's Two Main Theorems and Comments

Let $\mathcal{G}$ be a given family of distributions for Λ .

Definition: For each $G_o \in \mathcal{G}$ let $\delta_{G_o}(X)$ be a solution, if one exists, such that

$$\int_{\mathcal{X}} \int_{\omega} \delta_{G_o} f(x|\lambda) dG_o d\mu \leq \alpha \qquad \text{(a)}$$

and δ_{G_o} maximizes (3)

$$\int_{\mathcal{X}} \int_{\omega^c} \delta f(x|\lambda) dG d\mu \quad \forall G \in \mathcal{G} \text{ and } \delta \in D^{\alpha}_{G_o} \qquad \text{(b)}$$

where $D^{\alpha}_{G_o} = \{\delta : \int_{\mathcal{X}} \int_{\omega} \delta f(x|\lambda) dG_o d\mu \leq \alpha\}$. Then his next theorem states:

Theorem 2.3.1. If G^* is a distribution such that δ_{G^*} exists and $\delta_{G^*} \in D^{\alpha}_{\mathcal{G}}$ then

(i) δ_{G^*} is a solution of (2) relative to $\mathcal{G}$.

(ii) If δ_G exists for each $G \in \mathcal{G}$ then

$$\int_{\mathcal{X}} \int_{\omega^c} \delta_{G^*} f(x|\lambda) dG d\mu \leq \int_{\mathcal{X}} \int_{\omega^c} \delta_G f dG d\mu \quad \forall G \in \mathcal{G}$$

i.e. G^* is a least favourable distribution. To prove (i) we note that 2(a) holds. Next $\delta \in D^{\alpha}_{\mathcal{G}} \Longrightarrow \delta \in D^{\alpha}_{G^*}$ hence 3(b) and therefore 2(b) holds. To prove (ii) we note that $\delta_{G^*} \in D^{\alpha}_{\mathcal{G}} \Longrightarrow \delta_{G^*} \in D^{\alpha}_{G}$ $\forall G \in \mathcal{G}$. Hence G^* is a least favourable distribution. □

He then goes on to use this theorem to prove theorem 2.3.2.

He indicates Theorem 2.3.1 may be useful in investigating problems where UMP Bayesian tests do not exist or problems are being considered that are not testing problems. This may be so but in his Chapter III, he has not used this theorem anywhere else, and as will be observed there is no need to use Theorem 2.3.1 to prove Theorem 2.3.2.

Theorem 2.3.2. Let X and Λ be real-valued random variables and assume

(i) $f(x|\lambda)$ has a monotone likelihood ratio in x, i.e. for $\lambda' > \lambda$ the ratio $\dfrac{f(x|\lambda')}{f(x|\lambda)}$ is a nondecreasing function of x.

(ii) $H : \omega = (\lambda \leq \lambda_o)$ against $A : \omega^c = (\lambda > \lambda_o)$. Then for each $G \in \mathcal{G}$

$$\begin{aligned} \delta_G(x) &= 1 \quad \text{if} \quad x > c_G \\ &= \gamma_G \quad \text{if} \quad x = c_G \\ &= 0 \quad \text{if} \quad x < c_G . \end{aligned} \tag{4}$$

And if $\exists$ a $G^* \in \mathcal{G} \ni$

$$\delta_{G^*}(x) = \inf_{G \in \mathcal{G}} \delta_G(x) \quad \forall x$$

then δ_{G^*} is a solution of (2) relative to $\mathcal{G}$. If $\forall G \in \mathcal{G}$ there is a solution to (3), then G^* is a least favourable distribution.

Proof: That δ_G is given by (4) follows from the fact that for the MLR family and H against A as in (ii), $\forall \lambda < \lambda_o$ the UMP classical

test also minimizes error of type I subject to $\int_{\mathscr{X}} \delta f(x|\lambda_o)d\mu = \alpha$.

Next $\delta_{G^*}(X) = \inf_{G\varepsilon\mathscr{G}} \delta_G(X) \Longrightarrow \delta_{G^*}(X) \varepsilon D^{\alpha}_{\mathscr{G}}$. Let $\delta \varepsilon D^{\alpha}_{\mathscr{G}}$. Without loss of generality assume

$$\begin{aligned}\delta(x) &= 1 \quad \text{if} \quad x > c \\ &= \gamma \quad \text{if} \quad x = c \quad 0 \leq \gamma \leq 1 \\ &= 0 \quad \text{if} \quad x < c\,.\end{aligned}$$

Then

$$\begin{aligned}\delta \varepsilon D^{\alpha}_{\mathscr{G}} &\Longrightarrow \int_{\mathscr{X}} \int_{\omega} \delta f(x|\lambda)dG^* d\mu \leq \alpha \\ &\Longrightarrow \int_{\mathscr{X}} \int_{\omega} [\delta-\delta_{G^*}]f(x|\lambda)dG^* d\mu \leq 0 \\ &\Longrightarrow \delta_{G^*} \geq \delta \quad \text{a.e.} \quad \mu \\ &\Longrightarrow \int_{\mathscr{X}} \int_{\omega^c} [\delta_{G^*}-\delta]f(x|\lambda)dGd\mu \geq 0 \quad \forall G \varepsilon \mathscr{G}. \\ &\Longrightarrow 2(b)\,. \quad \square\end{aligned}$$

There is thus no need to appeal to Theorem 2.3.1.

Having proved Theorem 2.3.2 he has given four examples where such tests exist. In these he has verified the existence of a $G^* \varepsilon \mathscr{G}$ such that

$$\delta_{G^*}(X) = \inf_{G\varepsilon\mathscr{G}} \delta_G(X)$$

and that δ_{G^*} is not trivially zero or one everywhere. Only in two

examples has he attempted to construct the corresponding tests; in one on the basis of two unproved statements. He also suggests an approximate construction, which can be used when these statements are not true. When Λ is normal he has not mentioned construction at all and it does seem that exact construction is possible in this case; this will be attempted in Chapter III, after considering a result proved by assuming that there is a $G^* \varepsilon \mathcal{G}$ such that G^* is a least favourable distribution. According to this result the construction of the UMP Bayesian tests in the examples considered is possible whenever we know a G^* which is least favourable.

CHAPTER III

A CHARACTERIZATION OF THE LEAST FAVOURABLE DISTRIBUTION, AND THE LOCATION OF θ^*

3.1 On Using the Concept of a Least Favourable Distribution

Let $\mathcal{G} = \{G_\theta : G_\theta$ is the c.d.f. of Λ , $\theta \varepsilon \Theta\}$ be a parametric family of distributions. Then in certain families and for certain hypothesis-testing problems it seems, intuitively, that guessing the least favourable distribution should not be difficult. In such families one may guess the least favourable distribution from the fact that such a distribution renders least help in deciding whether $\lambda \varepsilon \omega$ or $\lambda \varepsilon \omega^c$.

For example let $\mathcal{G} = \{G_\theta : \Lambda$ is $N(\theta,1)$, $-\infty < \theta < \infty\}$ and $f(x|\lambda) = N(\lambda,1)$; further let $H : \lambda \varepsilon \omega = (\lambda \leq \lambda_o)$ against $A : \lambda \varepsilon \omega^c = (\lambda > \lambda_o)$. Then for usual α and $\forall\theta$ such that $|\theta - \lambda_o| \geq 3$ the decision making is made very easy: either always accept, or always reject, according as $\theta < \lambda_o$ or $\theta > \lambda_o$. However when $\theta = \lambda_o$ $G_{\lambda_o}(\omega) = G_{\lambda_o}(\omega^c) = \frac{1}{2}$ and the distribution is the least favourable. For any other value of θ , $G_\theta(\omega) \neq \frac{1}{2}$ and the distribution G_θ is of more help to the statistician than G_{λ_o} .

Now consider the special case $\lambda_o = 0$ with $\mathcal{G}$ and $f(x|\lambda)$ as specified in the previous paragraph. For this case Meeden has proved that the cut-off point c_θ is such that

(i) $\lim_{\theta \to +\infty} c_\theta = -\infty$

(ii) $\lim_{\theta \to -\infty} c_\theta = -\infty$

(iii) $\exists$ a$\theta \ni c_\theta > -\infty$

(iv) c_θ is a continuous function of θ , where c_θ is given by

$$\int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x|\lambda)\, dG_\theta\, d\mu \leq \alpha$$

If our intuitive guess is correct, then, by Meeden's theorem

$$\delta_{\theta=0}(x) = \inf_{G\varepsilon \mathcal{G}} \delta_G(x) = \inf_{\theta\varepsilon R} \delta_\theta(x) .$$

Since for any θ $\delta_\theta(x)$ is

$$\begin{aligned} \delta_\theta(x) &= 1 \quad x \geq c_\theta \\ &= 0 \quad x < c_\theta \end{aligned}$$

$$\delta_{\theta=0}(x) = \inf_{\theta\varepsilon R} \delta_\theta(x) \iff c_0 = \sup_{\theta\varepsilon\Theta} c_\theta .$$

$\Longrightarrow$) $\delta_{\theta=0}(c_0) = 1 \Longrightarrow \inf_{\theta\varepsilon R} \delta_\theta(c_0) = 1 \Longrightarrow \delta_\theta(c_0) = 1 \quad \forall\theta \;\varepsilon\; R$

$\Longrightarrow c_0 \geq c_\theta \quad \forall\theta \;\varepsilon\; R \Longrightarrow c_0 \geq \sup_{\theta\varepsilon R} c_\theta$. Clearly $c_0 \leq \sup_{\theta\varepsilon R} c_\theta$. Hence $c_0 = \sup_{\theta\varepsilon R} c_\theta$.

$\Longleftarrow$) Suppose $x \geq c_0$ then $\delta_{\theta=0}(x) = 1$; also $c_0 \geq c_\theta \quad \forall\theta$ $\Longrightarrow \delta_\theta(x) = 1 \quad \forall\theta \;\varepsilon\; R$. If $x < c_0$ then $\delta_{\theta=0}(x) = 0$ $\Longrightarrow \delta_{\theta=0}(x) \leq \delta_\theta(x) \quad \forall\theta \Longrightarrow \delta_{\theta=0}(x) \leq \inf_{\theta\varepsilon R} \delta_\theta(x)$ hence $\forall x \quad \delta_{\theta=0}(x) \leq \inf_{\theta\varepsilon R} \delta_\theta(x)$. Clearly $\delta_{\theta=0}(x) \geq \inf_{\theta\varepsilon R} \delta_\theta(x)$. Hence $\delta_{\theta=0}(x) = \inf_{\theta\varepsilon R} \delta_\theta(x)$.

Hence if our guess is correct $\forall \alpha\ 0 < \alpha < 1$, then c_θ should attain its sup at $\theta = 0$. Now consider $\alpha = \frac{1}{2}$; then c_o is $\ni$

$$\int_{c_o}^{+\infty} \int_{-\infty}^{o} f(x|\lambda) dG_o d\mu \leq \frac{1}{2} .$$

Here equlity is possible when $c_o = -\infty$. For then

$$\int_{-\infty}^{o} \int_{-\infty}^{+\infty} f(x|\lambda) d\mu dG_o = \int_{-\infty}^{o} dG = \frac{1}{2} .$$

On the other hand, by Meeden's result, $\forall\ \alpha\ 0 < \alpha < 1$, $\exists\ a\theta \in R$ $\ni c_\theta > -\infty$ hence $c_o \neq \sup_{\theta\varepsilon\Theta} c_\theta$ and as guessed G_o is not the least favourable distribution. Hence it is not true that G_o is the least favourable distribution for all α . It appears that θ^* is a function of α . Computations in the form of plots of θ against c_θ , for $\alpha = .005$ and $\alpha = .05$, also suggest θ^* as a function of α .

Next if we consider $f(x|\lambda) = N(\lambda,\tau^2)$ and $f(\lambda) = N(\theta,\sigma^2)$ then it seems that θ^* is a function of (α,τ,σ). In Section 3.3 we shall derive an equation in c_{θ^*}, θ^*, τ and σ for G_{θ^*} to be the least favourable distribution.

In next section a characterization of G_{θ^*} is obtained. Let $\mathcal{G}$ denote a parametric family; then for each $\theta \in \Theta$, by the Neymann-Pearson lemma there exists a most powerful Bayesian test of $H : \lambda \in \omega = (\lambda \leq \lambda_o)$ against $A : \lambda \in \omega^c = (\lambda > \lambda_o)$. In such a family we say that a distribution $G_{\theta^*} \in \mathcal{G}$ is least favourable iff the most powerful test $\delta_{G_{\theta^*}}$ of H against A satisfies

$$\int_{\mathscr{X}} \int_{\omega^c} \delta_{G_{\theta^*}} f(x|\lambda)\, dG_{\theta^*} d\mu \le \int_{\mathscr{X}} \int_{\omega^c} \delta_{G_\theta} f(x|\lambda)\, dG_\theta d\mu \quad \forall\theta \in \Theta$$

3.2 A Characterization of the Least Favourable Distribution

Theorem 3.2.1. Let

(i) $\mathscr{G} = \{G_\theta : \theta\varepsilon\Theta\}$

(ii) $f(x|\lambda)$ be a monotone likelihood ratio family in x ; i.e., for $\lambda' > \lambda$, $\dfrac{f(x|\lambda')}{f(x|\lambda)}$ is a nondecreasing function of x .

(iii) $H : \lambda \varepsilon \omega = (\lambda \le \lambda_o)$ against $A : \lambda \varepsilon \omega^c = (\lambda > \lambda_o)$. Then G_{θ^*} is the least favourable distribution iff $\delta_{\theta^*}(x) \varepsilon D^{\alpha}_{\mathscr{G}}$, where $\delta_{\theta^*} \equiv \delta_{G_{\theta^*}}$.

Proof: Let G_{θ^*} be the least favourable distribution. As seen in Chapter II $\forall\theta \varepsilon \Theta$

$$\begin{aligned} \delta(x) &= 1 \quad & x > c_\theta & \\ &= \gamma_\theta \quad & x = c_\theta & \quad (0 \le \gamma_\theta \le 1) \\ &= 0 \quad & x < c_\theta & . \end{aligned}$$

Since G_{θ^*} is the least favourable distribution

$$\int_{\omega^c} \int_{\mathscr{X}} [\delta_{\theta^*} - \delta_\theta] f(x|\lambda)\, dG_\theta d\mu \le 0 \quad \forall\theta \varepsilon \Theta \tag{1}$$

For a given θ let

$$S^+ = \{x : \delta_{\theta^*} - \delta_\theta > 0\}$$

$$\subset \{x: x \geq c_{\theta^*},\ x \leq c_\theta\} = \{x: c_{\theta^*} \leq x \leq c_\theta\} = S \quad \text{say}$$

then on S^c $\delta_{\theta^*} - \delta_\theta = 0$. Suppose $\mu(S) > 0$ then $\int_{\omega^c} \int_{\mathfrak{X}} [\delta_{\theta^*} - \delta_\theta] f d\mu dG_\theta > 0$ assuming $G(\omega^c) > 0$, i.e. a contradiction to (1) above. Hence $\delta_{\theta^*} \leq \delta_\theta$ a.e. μ $\forall\theta \in \Theta$. Hence $\delta_{\theta^*} \in D^\alpha_{\mathcal{G}}$. Now suppose $\delta_{\theta^*} \in D^\alpha_{\mathcal{G}}$ then $\delta_{\theta^*} \in D^\alpha_G$ $\forall G \in \mathcal{G}$. Since δ_θ is the MP Bayesian test we have

$$\int_{\omega^c} \int_{\mathfrak{X}} [\delta_{\theta^*} - \delta_\theta] f(x|\lambda) d\mu dG_\theta \leq 0 \tag{1a}$$

using S^+ and S as above, we have $\delta_{\theta^*} - \delta_\theta = 0$ on S^c as before; and if $\mu(S) > 0$ then we have

$$\int_{\omega^c} \int_{\mathfrak{X}} [\delta_{\theta^*} - \delta_\theta] f d\mu dG_\theta > 0 \quad \text{assuming} \quad G(\omega^c) > 0$$

i.e. a contradiction to (1a) above. Hence $\delta_{\theta^*} \leq \delta_\theta$ a.e. μ $\forall\theta \in \Theta$. Hence G_* is the least favourable distribution. □

Corollary 3.2.1. If conditions (i)-(iii) of the above theorem are satisfied and if for some $\theta^* \in \Theta$ G_{θ^*} is the least favourable distribution then

(i) $\delta_{\theta^*}(X)$ is a solution of (2), as in Section 2.1.2.

(ii) $\delta_{\theta^*}(x) = \inf_{\theta \in \Theta} \delta_\theta(x) \quad \forall x$

Proof: (i) By the above theorem $\delta_{\theta^*}(x) \in D^\alpha_{\mathcal{G}}$; hence (2a). Also by the same theorem $\delta_{\theta^*}(x) \leq \delta_\theta(x)$ $\forall\theta \in \Theta$. Hence $\delta_{\theta^*}(x)$ is a

solution of (2).

(ii) $\delta_{\theta^*}(x) \leq \delta_\theta(x) \quad \forall\theta \in \Theta \Longrightarrow \delta_{\theta^*}(x) \leq \inf_{\theta\in\Theta} \delta_\theta(x)$, and clearly $\delta_{\theta^*}(x) \geq \inf_{\theta\in\Theta} \delta_\theta(x)$. Hence $\delta_{\theta^*}(x) = \inf_{\theta\in\Theta} \delta_\theta(x)$.

Corollary 3.2.2. If

(i) The conditions of corollary 3.2.1 are satisfied;

(ii) Θ is a subset of a Euclidian Space;

(iii) $S = \sup_{\theta\in\Theta} c_\theta$ is attained at a point $\theta' \in \Theta$, where θ' is a limit point of Θ, then $\theta' = \theta^*$ or $c_{\theta^*} = \sup_{\theta\in\Theta} c_\theta$.

Proof: Since $S = \sup_{\theta\in\Theta} c_\theta$ is attained at θ', $S = c_{\theta'}$. Clearly $c_{\theta'} \geq c_{\theta^*}$. Let $x < c_{\theta'}$; since θ' is a limit point of Θ, $\exists$ a $\theta \in \Theta \ni c_{\theta'} > c_\theta > x$; but $c_\theta > x \Longrightarrow \delta_\theta(x) = 0 \Longrightarrow \delta_{\theta^*}(x) = \inf_{\theta\in\Theta} \delta_\theta(x)$ $\inf_{\theta\in\Theta} \delta_\theta(x) = 0 \Longrightarrow x < c_{\theta^*}$. Since this is true for all $x < c_{\theta'}$, however close to $c_{\theta'}$, we have $c_{\theta'} \leq c_{\theta^*}$. Hence $c_{\theta'} = c_{\theta^*}$ or $\theta' = \theta^*$. □

A particular case of this corollary, when $\mathcal{G} = \{G_\theta : \Lambda = N(\theta,1), -\infty<\theta<\infty\}$ was proved in Section 3.1. In this case, the fact that $c_{\theta^*} = \sup_{\theta\in\Theta} c_\theta$, is useful in finding θ^*, and hence constructing $\delta_{\theta^*}(x)$, which Meeden has not done. Similarly, it seems, that Corollary 3.2.2 will be useful in constructing UMP Bayesian tests in other examples considered by him.

In the following section we give a heuristic proof of the piecewise monotonicity of c_θ in the normal case, then derive an equation in θ^* and other variables.

3.3 Piecewise Monotonicity of c_θ

$$\text{Let } \mathcal{G} = \{G_\theta : \Lambda = N(\theta, \tau^2),\ -\infty < \theta < +\infty,\ \tau^2 > 0\}$$

$$f(x|\lambda) = N(\lambda, \sigma^2)$$

$$H : \lambda \leq 0 \quad \text{against} \quad A: \lambda > 0 .$$

For the construction of the UMP Bayesian test we look for the least favourable distribution, i.e. for θ^* such that G_{θ^*} is the least favourable distribution. For a$\theta \in R$ c_θ is defined as follows

$$\delta_\theta(x) = 1 \quad \text{if} \quad x \geq c_\theta$$
$$= 0 \qquad x < c_\theta$$

such that

$$\int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x,\lambda)\, d\lambda dx \leq \alpha, \quad 0 < \alpha < 1 \tag{2}$$

Clearly $c_\theta = -\infty$ if $\int_{-\infty}^{0} f_\theta(\lambda)\, d\lambda \leq \alpha$ and we require that $\int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x,\lambda)\, d\lambda dx$ should be as large as possible. If $\int_{-\infty}^{0} f_\theta(\lambda)\, d\lambda > \alpha$ then

$$\int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x,\lambda)\, d\lambda dx = \alpha \quad \text{with} \quad c_\theta > -\infty \tag{3}$$

Now for a given

$$f_\theta(x,\lambda) = \frac{1}{2\pi\tau\sigma} \exp[-\frac{1}{2}\{(\frac{x-\lambda}{\sigma})^2 + (\frac{\lambda-\theta}{\tau})^2\}]$$

writing $(x-\lambda)^2 = [(x-\theta)-(\lambda-\theta)]^2$ we have

$$f_\theta(x,\lambda) = \frac{1}{2\pi\tau\sigma}\exp[-\frac{1}{2}\{\frac{(x-\theta)^2}{\sigma^2} - \frac{2}{\sigma^2}(x-\theta)(\lambda-\theta)+(\frac{1}{\sigma^2}+\frac{1}{\tau^2})(\lambda-\theta)^2\}]$$

$$= \frac{1}{2\pi\tau\sigma}\exp[-\frac{1}{2}\frac{(\tau^2+\sigma^2)}{\sigma^2}\{\frac{(x-\theta)^2}{(\tau^2+\sigma^2)} - \frac{2\tau}{\sqrt{\tau^2+\sigma^2}}\cdot\frac{(x-\theta)}{\sqrt{\tau^2+\sigma^2}}\cdot\frac{(\lambda-\theta)}{\tau}+\frac{(\lambda-\theta)^2}{\tau^2}\}]$$

i.e. for a given θ (x,λ) is a bivariate normal vector with common mean θ, var $X = (\tau^2+\sigma^2)$, var $\Lambda = \tau^2$ and correlation coefficient $\tau/\sqrt{\tau^2+\sigma^2}$.

Hence, for $\theta << 0$ most of the probability mass is distributed over the third quadrant; as θ increases the common mean, θ , shifts along the line $x = \lambda$, and the probability mass over the first quadrant increases.

On the other hand the mass, $\int_k^\infty \int_{-\infty}^0 f(x,\lambda)d\lambda dx$, over the shaded region of Figure I increases as θ increases to a value θ^* and then it decreases. For large and negative θ ,

$$\int_{-\infty}^0 f(x,\lambda)d\lambda \simeq f(x) = N(\theta,(\tau^2+\sigma^2))$$

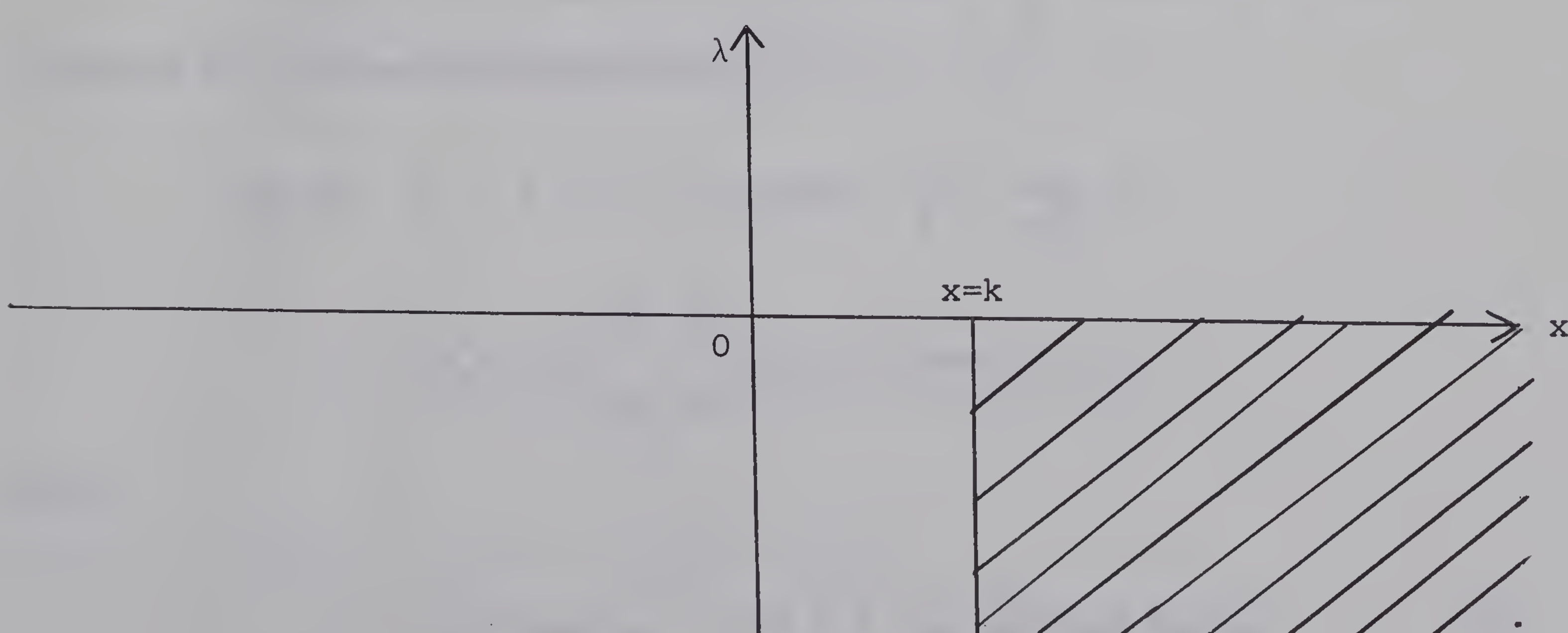


FIGURE I

and hence as θ increases c_θ must increase for

$\int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x,\lambda)d\lambda dx = \alpha$ to hold $(\theta << 0)$. Moreover, if $\theta_2 > \theta_1$, and if

$$\int_{c_{\theta_1}}^{+\infty} \int_{-\infty}^{0} f_{\theta_1}(x,\lambda)d\lambda dx = \alpha \quad \text{and} \quad \int_{c_{\theta_1}}^{+\infty} \int_{-\infty}^{0} f_{\theta_2}(x,\lambda)d\lambda dx > \alpha$$

then for $\int_{c_{\theta_2}}^{+\infty} \int_{-\infty}^{0} f_{\theta_2}(x,\lambda)d\lambda dx = \alpha$ to hold c_{θ_2} must be greater than c_{θ_1}. And it seems that c_θ increases as θ increases to a value θ^* and then c_θ decreases as θ increases.

Hence we take c_θ to be piecewise monotone and therefore differentiable almost everywhere on $\Theta = \{\theta : \int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x,\lambda)d\lambda dx = \alpha, c_\theta > -\infty\}$. On Θ c_θ is sufficiently smooth and we assume it differentiable everywhere on Θ. From Sections 3.1 and 3.2 we know that $c_{\theta^*} = \sup_{\theta \varepsilon \Theta} c_\theta$ since c_θ is differentiable on Θ we obtain θ^* by differentiating and solving $\frac{dc_\theta}{d\theta} = 0$ for θ^*.

It should be remarked that plots of c_θ against θ also indicate the piecewise monotonicity of c_θ.

For all α, $0 < \alpha < 1$, since $c_{\theta^*} = \sup_{\theta \varepsilon \Theta} c_\theta$,

$$\theta^* \varepsilon \Theta = \{\theta : \int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x,\lambda)d\lambda dx = \alpha, c_\theta > -\infty\} .$$

Also

$$\int_{c_\theta}^{+\infty} \int_{-\infty}^{0} f(x,\lambda)d\lambda dx = \alpha \quad \forall \theta \varepsilon \Theta \text{ as defined above.} \tag{4}$$

and we have, in the following lemma, an equation to locate θ^*.

Lemma 3.3.1 θ^* is the solution of the equation

$$\exp[-\frac{1}{2}\frac{\theta^2}{\tau^2}]\Phi\{-\frac{c_\theta}{\sigma}\}$$

$$= \frac{\tau}{\sqrt{\tau^2+\sigma^2}}\exp[-\frac{1}{2}(\frac{\theta^2}{\tau^2}+\frac{c_\theta^2}{\sigma^2})+\frac{1}{2}\frac{(\theta\sigma^2+c_\theta\tau^2)^2}{(\tau^2+\sigma^2)\tau^2\sigma^2}]\Phi\{-\frac{(\theta\sigma^2+c_\theta\tau^2)}{\tau\sigma\sqrt{\tau^2+\sigma^2}}\}$$

obtained by differentiating equation (4) with respect to θ in the following form:

$$\frac{1}{2\pi\sigma\tau}\int_{c_\theta}^{+\infty}\int_{-\infty}^{0}\exp[-\frac{1}{2}\{\frac{(x-\lambda)^2}{\sigma^2}+\frac{(\lambda-\theta)^2}{\tau^2}\}]d\lambda dx = \alpha \qquad \forall\theta \in \Theta$$

Proof: Differentiating the above equation with respect to θ we have:

$$-c_\theta'\int_{-\infty}^{0}\exp[-\frac{1}{2}\{\frac{(c_\theta-\lambda)^2}{\sigma^2}+\frac{(\lambda-\theta)^2}{\tau^2}]d\lambda$$

$$+\int_{c_\theta}^{+\infty}\int_{-\infty}^{0}\frac{(\lambda-\theta)}{\tau^2}\exp[-\frac{1}{2}\{\frac{(x-\lambda)^2}{\sigma^2}+\frac{(\lambda-\theta)^2}{\tau^2}\}]d\lambda dx = 0$$

or $c_\theta' = 0$ iff

$$-\int_{c_\theta}^{+\infty}\int_{-\infty}^{0}\frac{(\lambda-\theta)}{\tau^2}\exp[-\frac{1}{2}\{\frac{(x-\lambda)^2}{\sigma^2}+\frac{(\lambda-\theta)^2}{\tau^2}\}]d\lambda dx = 0$$

writing $-\frac{(\lambda-\theta)}{\tau^2} = -\frac{(x-\lambda)}{\sigma^2}-\frac{(\lambda-\theta)}{\tau^2}+\frac{(x-\lambda)}{\sigma^2}$, $c_\theta' = 0$ iff

$$\int_{c_\theta}^{+\infty} \exp[-\tfrac{1}{2}\{\frac{x^2}{\sigma^2} + \frac{\theta^2}{\tau^2}\}]dx$$

$$-\int_{c_\theta}^{+\infty}\int_{-\infty}^{0} \frac{(x-\lambda)}{\sigma^2} \exp[-\tfrac{1}{2}\{\frac{(x-\lambda)^2}{\sigma^2} + \frac{(\lambda-\theta)^2}{\tau^2}\}]d\lambda dx = 0$$

i.e. iff
$$e^{-\frac{1}{2}\cdot\theta^2/\tau^2} \int_{c_\theta}^{+\infty} \exp[-\tfrac{1}{2}\frac{x^2}{\sigma^2}]dx$$

$$= \int_{c_\theta}^{+\infty}\int_{-\infty}^{0} \frac{(x-\lambda)}{\sigma^2} \exp[-\tfrac{1}{2}\{\frac{(x-\lambda)^2}{\sigma^2} + \frac{(\lambda-\theta)^2}{\tau^2}\}]d\lambda dx \quad .$$

i.e.
$$\sqrt{2\pi}\,\sigma \exp[-\tfrac{1}{2}\frac{\theta^2}{\tau^2}]\Phi\{-\frac{c_\theta}{\sigma}\}$$

$$= \int_{-\infty}^{0}\int_{c_\theta}^{+\infty} \exp[-\tfrac{1}{2}\frac{(\lambda-\theta)^2}{\tau^2}]\ \frac{(x-\lambda)}{\sigma^2} \exp[-\tfrac{1}{2}\frac{(x-\lambda)^2}{\sigma^2}]dxd\lambda$$

$$= \int_{-\infty}^{0} \exp[-\tfrac{1}{2}\frac{(\lambda-\theta)^2}{\tau^2}]\ \exp[-\tfrac{1}{2}\frac{(c_\theta-\lambda)^2}{\sigma^2}]d\lambda$$

$$= \exp[-\tfrac{1}{2}(\frac{\theta^2}{\tau^2} + \frac{c_\theta^2}{\sigma^2})]\int_{-\infty}^{0} \exp[-\tfrac{1}{2}\{(\frac{1}{\tau^2} + \frac{1}{\sigma^2})\lambda^2 - 2(\frac{\theta}{\tau^2} + \frac{c_\theta}{\sigma^2})\lambda\}]$$

$$= \exp[-\tfrac{1}{2}(\frac{\theta^2}{\tau^2} + \frac{c_\theta^2}{\sigma^2})]\int_{-\infty}^{0} \exp[-\tfrac{1}{2}\frac{(\tau^2+\sigma^2)}{\tau^2\sigma^2}\{\lambda^2 - 2\frac{(\theta\sigma^2+c_\theta\tau^2)}{(\tau^2+\sigma^2)}\lambda\}]d\lambda$$

$$= \exp[-\tfrac{1}{2}(\frac{\theta^2}{\tau^2} + \frac{c_\theta^2}{\sigma^2})]\int_{-\infty}^{0} \exp[-\tfrac{1}{2}\frac{(\tau^2+\sigma^2)}{\tau^2\sigma^2}\{\lambda - \frac{\theta\sigma^2+c_\theta\tau^2}{(\tau^2+\sigma^2)}\}^2]\exp[\tfrac{1}{2}\frac{(\theta\sigma^2+c_\theta\tau^2)^2}{(\tau^2+\sigma^2)\tau^2\sigma^2}]d\lambda$$

$$= \exp[-\tfrac{1}{2}(\frac{\theta}{\tau^2} + \frac{c_\theta^2}{\sigma^2}) + \tfrac{1}{2}\frac{(\theta\sigma^2+c_\theta\tau^2)^2}{(\tau^2+\sigma^2)\tau^2\sigma^2}]\ \frac{\tau\sigma}{\sqrt{\tau^2+\sigma^2}}\int_{-\infty}^{-(\theta\sigma^2+c_\theta\tau^2)/\tau\sigma\sqrt{\tau^2+\sigma^2}} \exp[-\tfrac{1}{2}\Lambda^2]d\Lambda$$

where $\Lambda = \frac{\sqrt{\tau^2+\sigma^2}}{\tau\sigma} [\lambda - \frac{\theta\sigma^2+c_\theta\tau^2}{(\tau^2+\sigma^2)}]$

$$= \sqrt{2\pi} \frac{\tau\sigma}{\sqrt{\tau^2+\sigma^2}} \exp[-\frac{1}{2}(\frac{\theta^2}{\tau^2}+\frac{c_\theta^2}{\sigma^2})+\frac{1}{2}\frac{(\theta\sigma^2+c_\theta\tau^2)^2}{(\tau^2+\sigma^2)\tau^2\sigma^2}] \times \Phi\{-\frac{(\theta\sigma^2+c_\theta\tau^2)}{\tau\sigma\sqrt{\tau^2+\sigma^2}}\}$$

Hence the result:

$$\exp[-\frac{1}{2}\frac{\theta^2}{\tau^2}]\Phi\{-\frac{c_\theta}{\sigma}\} = \frac{\tau}{\sqrt{\tau^2+\sigma^2}} \exp[-\frac{1}{2}(\frac{\theta^2}{\tau^2}+\frac{c_\theta^2}{\sigma^2})+\frac{1}{2}\frac{(\theta\sigma^2+c_\theta\tau^2)^2}{(\tau^2+\sigma^2)\tau^2\sigma^2}]$$
$$\times\Phi\{-\frac{\theta\sigma^2+c_\theta\tau^2}{\tau\sigma\sqrt{\tau^2+\sigma^2}}\}. \quad \square$$

Corollary: If $\sigma^2 = \tau^2 = 1$ then θ^* is the solution of

$$\exp[-\frac{1}{2}\theta^2]\Phi\{-c_\theta\} = \frac{1}{\sqrt{2}}\exp[-\frac{1}{4}(c_\theta-\theta)^2]\Phi\{-\frac{\theta+c_\theta}{\sqrt{2}}\} .$$

Hence we confirm the statements made in Section 3.1 that where $f(x|\lambda) = N(\lambda,\sigma^2)$ and $f(\lambda) = N(\theta,\tau^2)$ θ^* is a function of (α,τ,σ); and in particular if $\sigma^2 = \tau^2 = 1$ then θ^* is a function of α .

Following are tables for θ against c_θ for $\alpha = .005$ and $\alpha = .05$

TABLE II

(i) $\alpha = .005$

Table of Values of c_θ

θ		-2.5	-2.0	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0	2.5
c_θ		1.01	1.28	1.61	1.76	1.86	1.84	1.73	1.48	1.12	.47	-1.23

From the graph $\theta^* \doteq -.4 \quad c_{\theta^*} \doteq 1.9$

$$\sqrt{2}\,\exp[-\tfrac{1}{4}(.4)^2]\Phi\{-1.9\} \doteq 1.414214 \times .923116 \times .0287$$

$$\doteq .03785$$

$$\exp[-\tfrac{1}{4}(2.3)^2]\Phi\{-\tfrac{1.9}{\sqrt{2}}\} \doteq .03862$$

Difference $.03862 - .03785 = .0008$.

TABLE III

(ii) $\alpha = .05$

Table of Values of c_θ

θ	-.8	-.7	-.6	-.5	-.4	-.3	-.2	-.1	0	.1	.2	.3	.4
c_θ	.713	.728	.74	.76	.75	.75	.73	.72	.69	.67	.64	.58	.54

From the graph $\theta^* \doteq -.5$, $c_{\theta^*} \doteq .76$,

$$\sqrt{2}\,\exp[-\tfrac{1}{2}(.25)]\Phi\{-.76\} \doteq 1.414214 \times .882497 \times .2236$$

$$\doteq .279062$$

Case (i) $\alpha = .005$

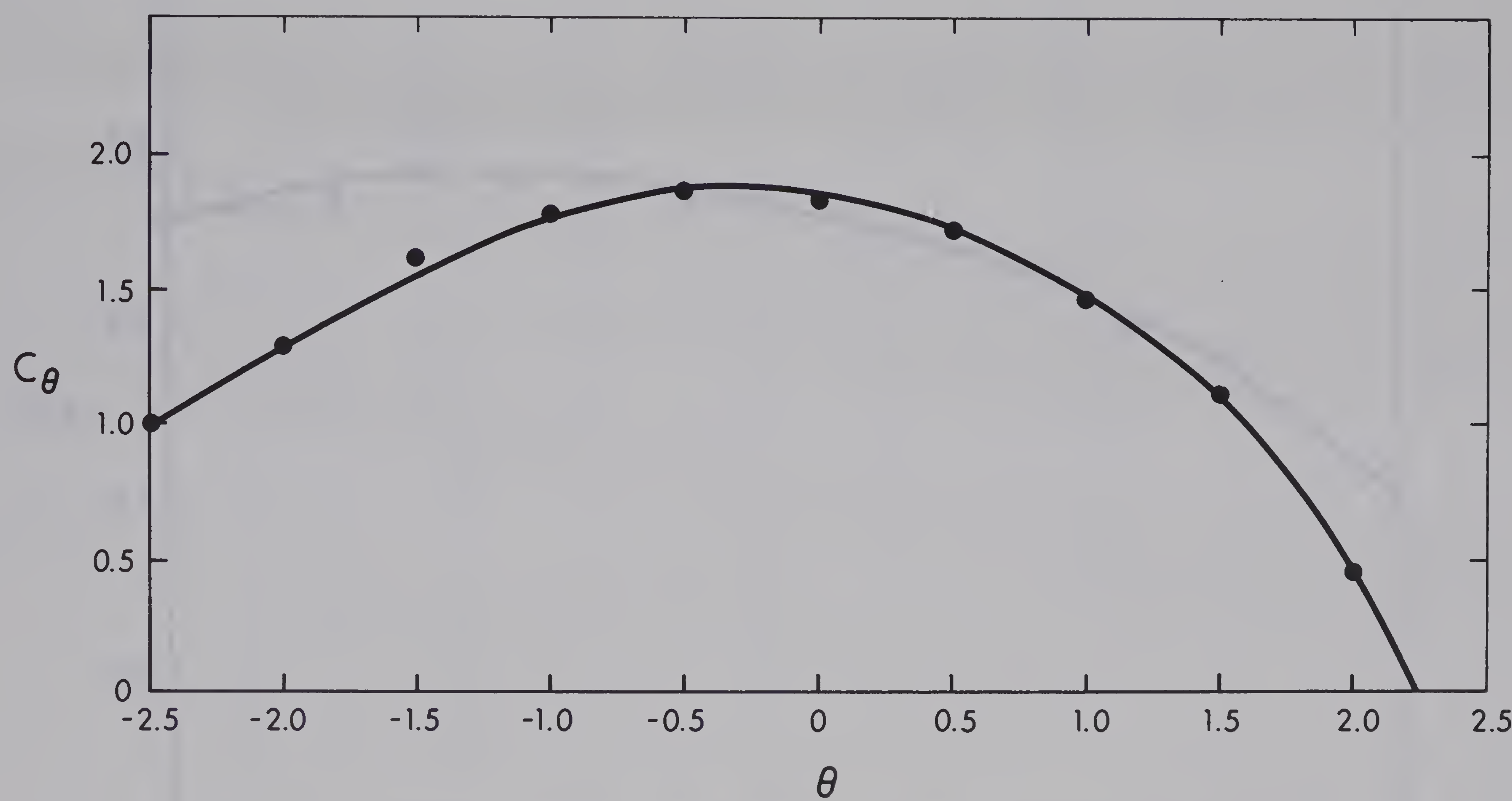
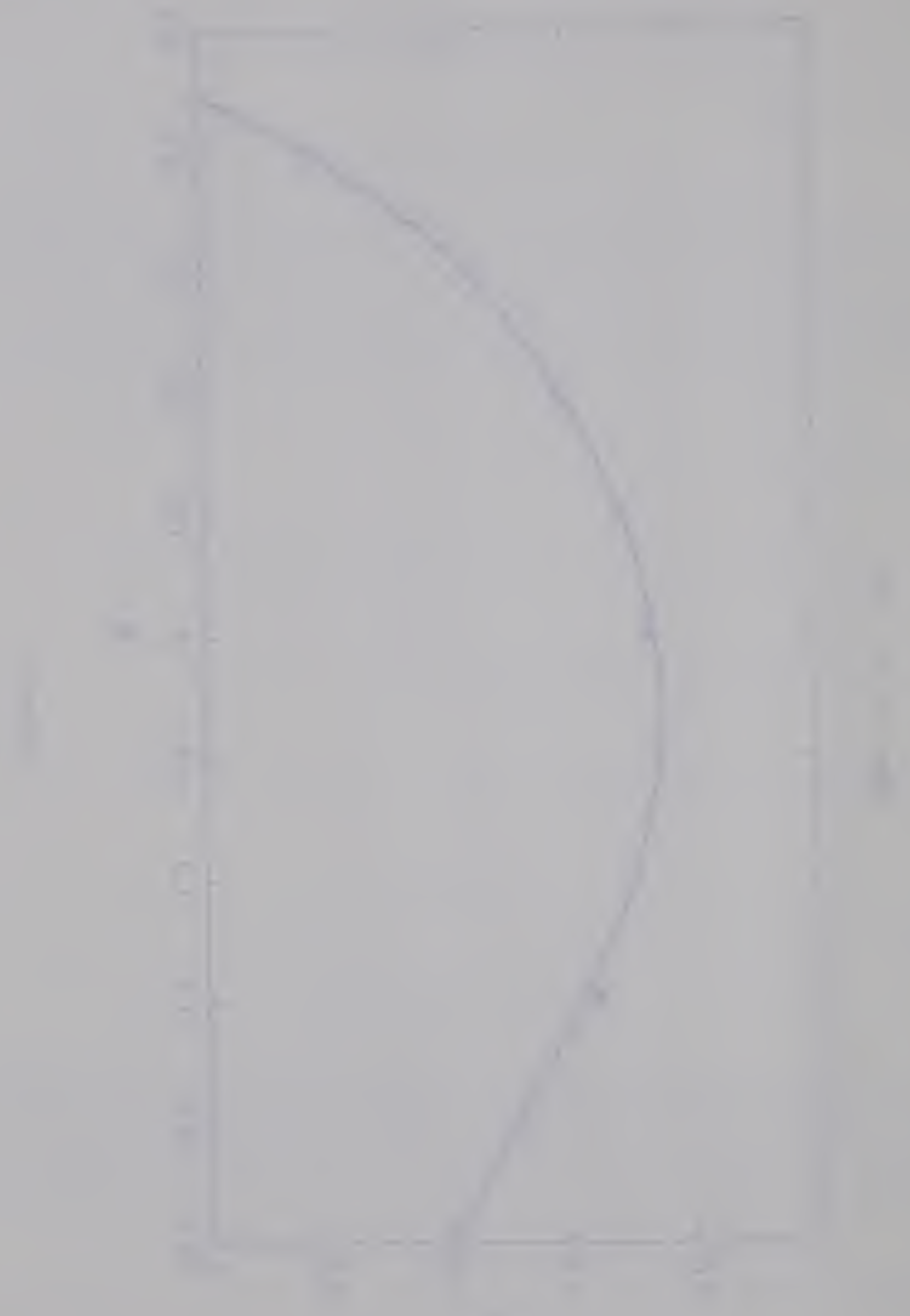


FIGURE II



Case (ii) $\alpha = .05$

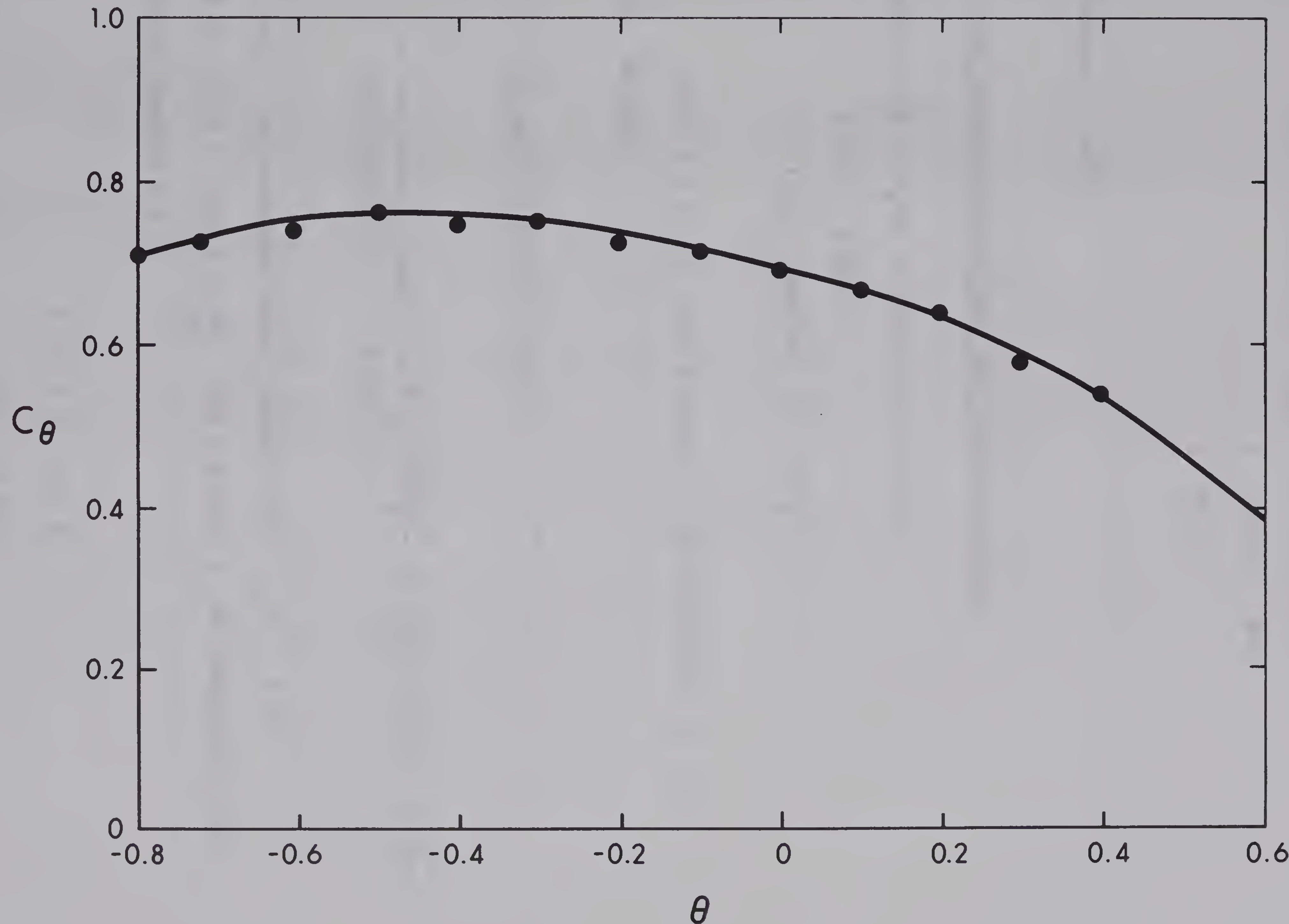
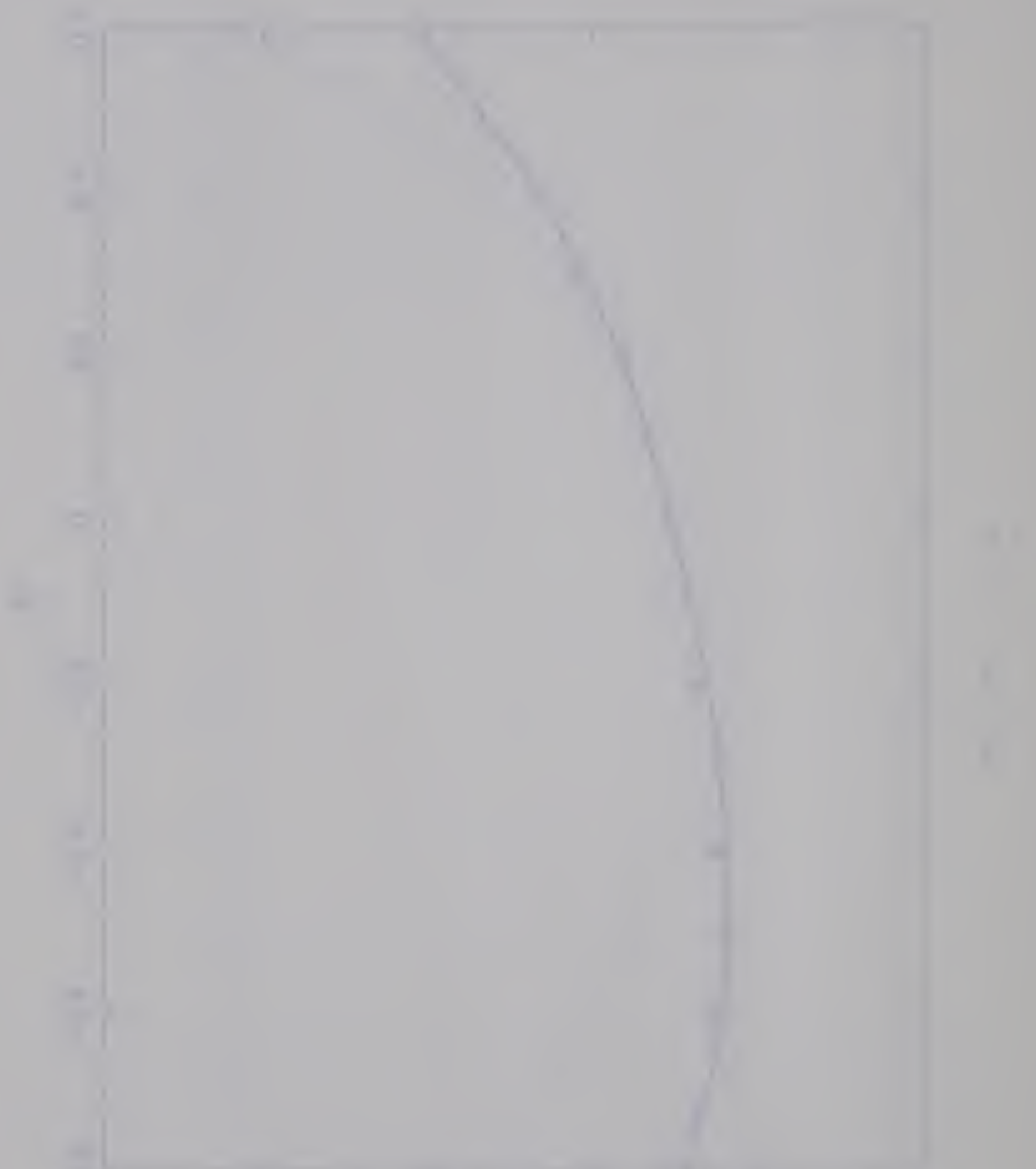


FIGURE III



$$\exp[-\frac{1}{4}(1.26)^2]\Phi\{-\frac{.26}{\sqrt{2}}\} \doteq \exp[.3969]\Phi\{-.1838\}$$

$$\doteq .672401 \times .4274$$

$$\doteq .28738$$

Difference $\doteq .00832$.

3.4 The Construction of the UMP Bayesian Tests

Example 1. $\mathcal{G} = \{G_\theta : \Lambda \text{ is } N(\theta,1), \ -\infty<\theta<+\infty\}$

$f(x|\lambda) = N(\theta,1)$

$H : (\lambda \le \lambda_o)$ against $A : (\lambda > \lambda_o)$.

Let $0 < \alpha < 1$. For a given θ to determine c_θ [or $\delta_\theta(x)$] we have:

$$f(x,\lambda) = \frac{1}{2\pi}\exp[-\frac{1}{2}\{(x-\lambda)^2+(\lambda-\theta)^2\}] \qquad \begin{matrix} -\infty < x < \infty \\ -\infty < \lambda < \infty \end{matrix}$$

$$= \frac{1}{2\pi\times\sqrt{2}\times 1\times\sqrt{1-\rho^2}}\exp[-\frac{1}{2(1-\rho^2)}\{\frac{(x-\theta)^2}{2} - 2\rho\frac{(x-\theta)}{\sqrt{2}}\frac{(\lambda-\theta)}{1} + \frac{(\lambda-\theta)^2}{1}\}]$$

$\therefore$ (X,Λ) has bivariate normal density with $\mu_x = \mu_\lambda = \theta$, $\sigma_x^2 = 2$, $\sigma_\lambda^2 = 1$ and $\rho = \frac{1}{\sqrt{2}}$. For a given θ as indicated in the proof of Theorem 2.3.2

$$\delta_\theta(x) = 1 \quad x \ge c_\theta$$

$$= 0 \quad x < c_\theta$$

where c_θ is determined by

$$\int_{c_\theta}^{+\infty} \int_{-\infty}^{\lambda_o} f(x,\lambda)\,dx\,d\lambda = \alpha$$

$f(x,\lambda)$ as in (5) above. To use the bivariate normal tables let

$$\frac{x-\theta}{\sqrt{2}} = x_1 \quad \text{and} \quad \frac{\lambda-\theta}{1} = -\lambda_1$$

then

$$\int_{(c_\theta-\theta)/\sqrt{2}}^{+\infty} \int_{\theta-\lambda_o}^{+\infty} f(x_1,\lambda_1)\,dx_1\,d\lambda_1 = \alpha$$

where $$f(x_1,\lambda_1) = \frac{1}{2\pi\times1\times1\sqrt{1-\frac{1}{2}}} \exp[-\{x_1^2+\lambda_1^2+2\times\frac{1}{\sqrt{2}}\, x_1\lambda_1\}] \tag{6}$$

i.e. bivariate normal with $\mu_{X_1} = \mu_{\lambda_1} = 0$, $\sigma^2_{X_1} = \sigma^2_{\lambda_1} = 1$ and $\rho = -\frac{1}{\sqrt{2}}$.

Letting $\lambda_o = 0$, $\alpha = .05$ the least favourable distribution corresponds to $\theta^* \simeq -.5$ hence $\delta_{\theta^*}(x)$ is the UMP Bayesian test and

$$\begin{aligned} \delta_{\theta^*}(x) &= 1 \quad \text{if} \quad x \geq c_{\theta^*} \\ &= 0 \qquad\quad\;\; < c_{\theta^*} \end{aligned}$$

where $\int_{(c_{\theta^*}+.5)/\sqrt{2}}^{+\infty} \int_{-.5}^{+\infty} f(x_1,\lambda_1)\,dx_1\,d\lambda_1 = .05$.

In order to use the standard bivariate integral tables the following relations are used.

$$L[h,k,\rho] = \int_{h}^{+\infty} \int_{k}^{+\infty} f(x,\lambda,\rho)\,dx\,d\lambda \quad h,k \geq 0 \tag{7}$$

where $f(x,\lambda,\rho) = \dfrac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\dfrac{1}{2(1-\rho^2)}\{x^2-2\rho x\lambda+\lambda^2\}\right]$

$$L[h,k,\rho] = L[k,h,\rho] \tag{8}$$

$$L[-h,k,\rho] = -L[h,k,-\rho] + \frac{1}{2}[1-\alpha(k)] \tag{9}$$

$$L[-h,-k,\rho] = L[h,k,\rho] + \frac{1}{2}\,[\alpha(h)+\alpha(k)] \tag{10}$$

where $\alpha(h) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-h}^{+h} \exp[-\frac{1}{2}\,x^2]\,dx$.

These are quoted from Tables of the Bivariate Normal Distribution Function and Related Functions.

Also (i) The approximation $\frac{1}{\sqrt{2}} \doteq .7$ rather than .71 is used.

(ii) Wherever necessary linear interpolation is used.

To compute the power of a test

$$\delta_\theta(x) = 1 \quad x \geq c_\theta$$

$$= 0 \quad < c_\theta$$

for $\alpha = .05$ and for a given value of θ we have, using the transformation $\frac{x-\theta}{\sqrt{2}} = x_1$ and $\lambda - \theta = \lambda_1$,

$$\text{Power} = \frac{1}{2\pi\sqrt{1-\frac{1}{2}}} \int_{(c_\theta-\theta)/\sqrt{2}}^{+\infty} \int_{-\theta}^{+\infty} \exp[-\frac{1}{2(1-\rho^2)} \{x_1^2-2\rho x_1\lambda_1+\lambda_1^2\} dx_1 d\lambda_1$$

$$\rho \doteq .7$$

Powers are computed for $\theta = \theta^*$, $\theta = 0$ and for the classical UMP test. The results are given in Table IV.

TABLE IV

Power of $\delta_{\theta*}(x)$, $\alpha = .05$

θ	-2.5	-2.0	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0	2.5
$\frac{.76-\theta}{\sqrt{2}}$	2.31	1.95	1.60	1.25	.89	.54	.18	-.17	-.52	-.88	-1.23
$p(\theta)$	.0019	.0079	.0248	.0638	.1380	.2452	.3373	.5497	.6994	.8069	.8897

Power of $\delta_o(x)$, $\alpha = .05$

θ	-2.5	-2.0	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0	2.5
$\frac{.69-\theta}{\sqrt{2}}$	2.26	1.90	1.55	1.20	.84	.49	.13	-.22	-.57	-.93	-1.28
$p(\theta)$	.0023	.0085	.0265	.0677	.1503	.2600	.4148	.5651	.7000	.8100	.8990

TABLE IV (Continued)
Power of $\delta_u(x)$ [Classical UMP], $\alpha = .05$

θ	-2.5	-2.0	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0	2.5
$\frac{1.65-\theta}{\sqrt{2}}$	2.93	2.58	2.23	1.87	1.52	1.17	.81	.46	.11	-.25	-.6
$P(\theta)$	.0007	.0025	.0085	.0230	.0560	.1220	.2030	.3290	.4550	.5980	.725

The power curves are plotted in Figure IV, we note that $\delta_{\theta^*}(x)$ and $\delta_o(x)$ have almost the same power curves and perhaps, for $\alpha = .05$, the guess that $\delta_o(x)$ is the UMP Bayesian test is not far off.

Example 2. $= \{f_\Lambda(\theta) = \binom{n}{\lambda}\theta^\lambda(1-\theta)^{n-\lambda} \quad \lambda = 0,1,2,\ldots,n,\ \theta\varepsilon[0,1]\}$

$$f(x|\lambda) = \frac{\binom{\lambda}{x}\binom{n-\lambda}{m-x}}{\binom{n}{m}} \quad \text{for} \quad x = 0,1,2,\ldots,m < n,$$

$$= 0 \quad \text{otherwise}$$

$$H : \lambda \leq \lambda_o \quad \text{against} \quad A : \lambda > \lambda_o .$$

We consider the case $n = 7$, $m = 6$, $\lambda_o = 5$ and $\alpha = .05$. Here θ^* is unknown yet but intuitively $\delta_{\theta=\theta_1}(x)$ seems close to $\delta_{\theta^*}(x)$ where θ_1 is such that $G_{\theta_1}(\omega) = \frac{1}{2}$.

To construct $\delta_{\theta_1}(x)$ we solve

$$h(\theta) = \sum_{\lambda=o}^{5} \binom{7}{\lambda}\theta^\lambda(1-\theta)^{7-\lambda} = \frac{1}{2}$$

Power Curves, $f(x|\lambda)$ normal, $\alpha = .05$

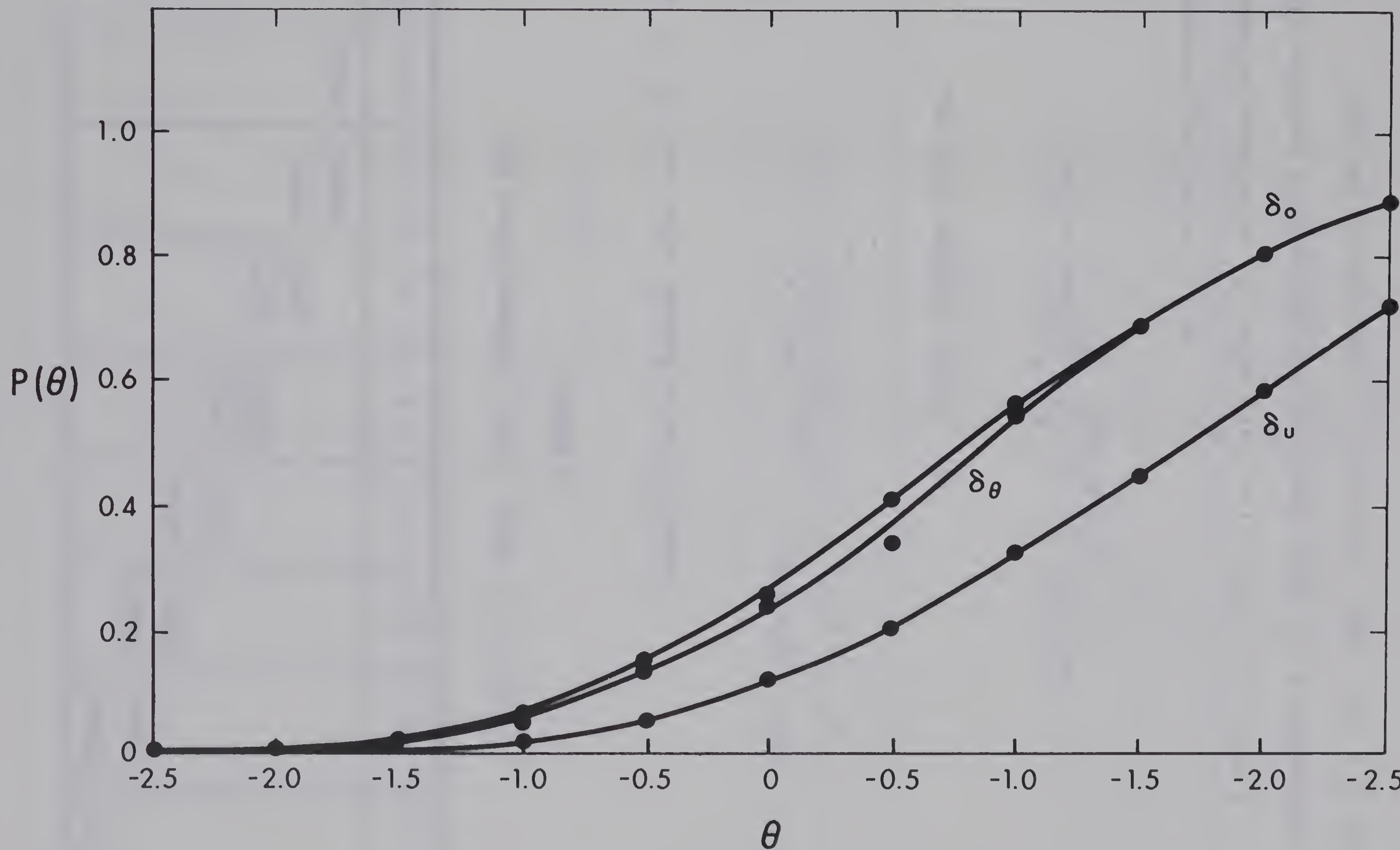
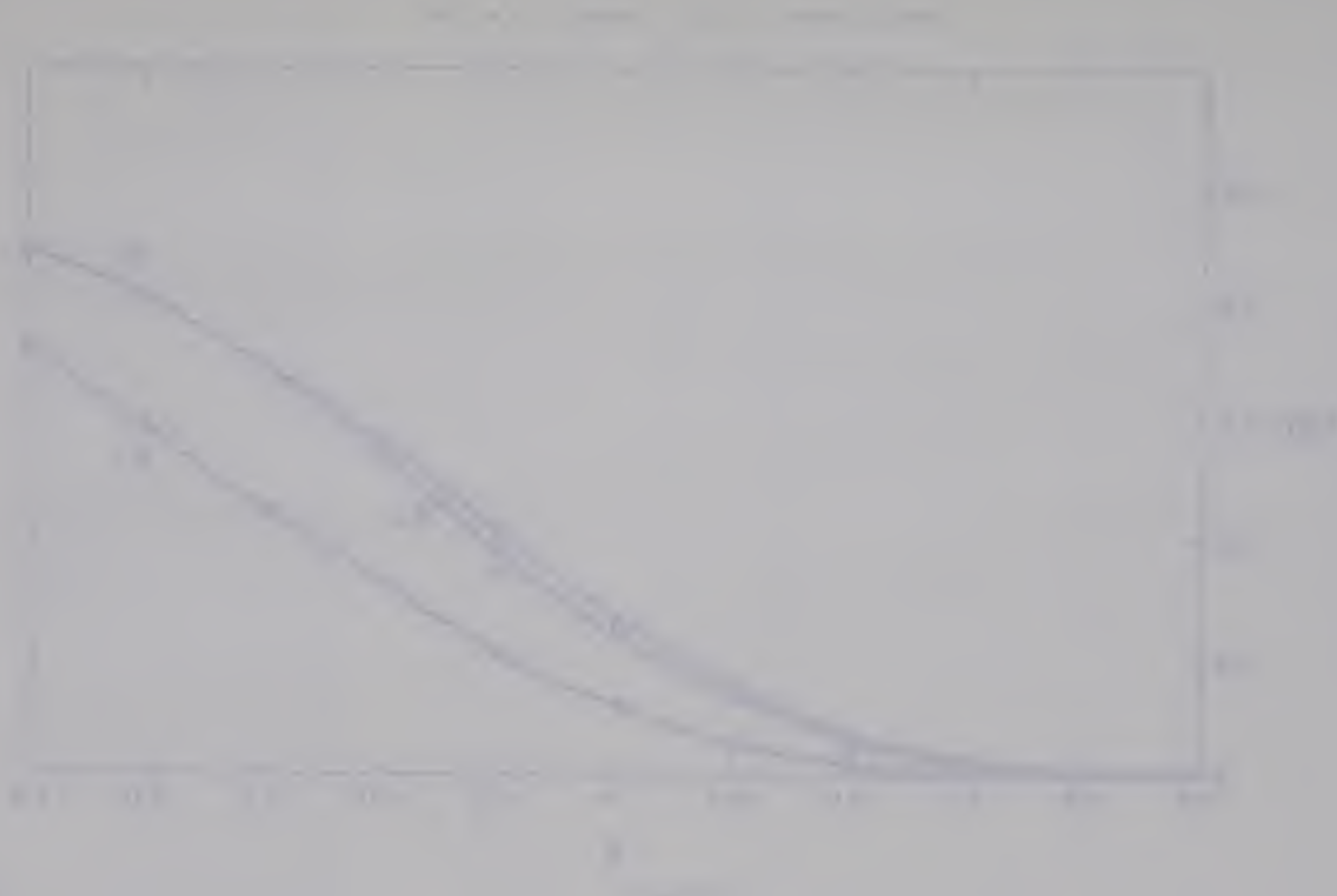


FIGURE IV



This has the solution $\theta_1 \simeq 0.77151$. That $h(\theta) = \frac{1}{2}$ has a unique solution follows from the fact that $h(\theta)$ is a continuous, decreasing function of θ, with $h(0) = 1$, $h(1) = 0$. It is a decreasing function of θ as

$$h(\theta) = (n-\lambda_o)\binom{n}{\lambda_o}\int_o^{1-\theta} t^{n-\lambda_o-1}(1-t)^{\lambda_o}dt .$$

The joint density $f(x,\lambda)$ is:

$$f(x,\lambda) = \binom{6}{x}\binom{1}{\lambda-x}\theta^{\lambda}(1-\theta)^{7-\lambda} \qquad \begin{array}{l} x = 0,1,\ldots,6 \\ \lambda = 0,1,\ldots,6,7 \end{array} \tag{11}$$

when $\theta = \theta_1$ the joint density is as in Table V.

TABLE V

The Joint Density $f(x,\lambda)$ When $\theta = .77151$

$x \backslash \lambda$	0	1	2	3	4	5	6	7
0	.00003	.00011						
1		.00066	.00222					
2			.00556	.01878				
3				.02503	.08453			
4					.06340	.21406		
5						.08562	.28911	
6							.04819	.16270

TABLE VI

Power of $\delta_{\theta_1}(x)$, $\alpha = .05$

θ	0	.2	.3	.4	.5	.6	.77151	.8	.9
$p(\theta)$	0	.0003	.0025	.0127	.0430	.1111	.3797	.4458	.7176

Power of $\delta_u(x)$, $\alpha = .05$

θ	0	.2	.3	.4	.5	.6	.77151	.8	.9
$p(\theta)$	.0000	.0001	.0013	.0067	.0238	.0663	.2615	.3151	.5872

Hence
$$\delta_{\theta_1}(x) = \begin{cases} 1 & x > 5 \\ \dfrac{.05}{.08562} & x = 5 \\ 0 & x < 5 \end{cases}$$

This follows from

$$\sum_{x=o}^{6} \sum_{\lambda=o}^{5} \delta_{\lambda_o}(x) f(x,\lambda) = .05$$

$f(x,\lambda)$ as in Table V. Power of $\delta_\theta(x)$ for any value of θ is given by

$$\sum_{x=o}^{6} \sum_{\lambda=6}^{7} \delta_\theta(x) f(x,\lambda) .$$

For $\theta = \theta_1$ this becomes

$$\frac{.05}{.08562}\{f(x=5,\lambda=6)+f(x=5,\lambda=7)\} + \{f(x=6,\lambda=6)+f(x=6,\lambda=7)\} = .3797 \,.$$

For some other values of θ, the power is given in Table VI.

Comparison with the classical UMP test, assuming λ = constant = 5 we have

$$f(x|5) = \frac{\binom{5}{x}\binom{2}{6-x}}{\binom{7}{6}} \qquad x = 0,1,2,\ldots,5$$

i.e. $f(x=4/5) = 5/7$ and $f(x=5|\lambda=5) = 2/7$. Hence

$$\begin{aligned}\delta_u(x) &= 1 && x > 5\\ &= \frac{.05}{2/7} = .175 && x = 5\\ &= 0 && \text{otherwise.}\end{aligned}$$

Power is tabulated in Table VI and curves are as shown in Figure V.

Conclusion: In conclusion, it has not been possible in this thesis to locate θ^* in other examples considered by Meeden. Even in the normal case the construction of $\delta_{\theta^*}(x)$ is carried out when θ^* is a function of α only. It might be of interest to study the relation between θ^* and τ , where Λ is $N(\theta,\tau^2)$ and $f(x|\lambda) = N(\lambda,1)$, when α is kept constant; perhaps $\lim_{\tau\to o} \theta^*$ exists.

Power Curves, $f(x|\lambda) = \binom{\lambda}{x}\binom{n-\lambda}{m-x}/\binom{n}{m}$, $\alpha = .05$

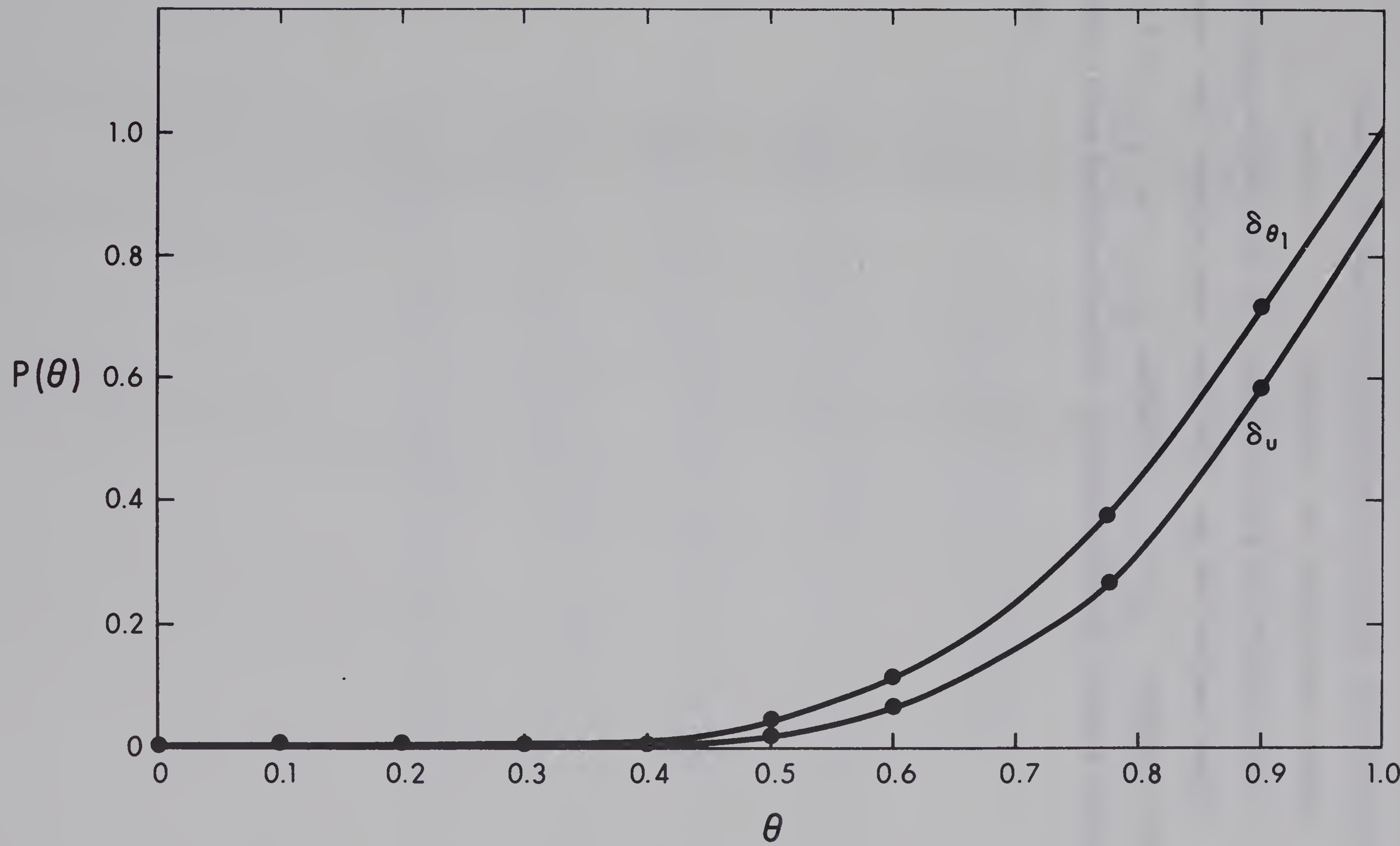


FIGURE V

Finally in case of the monotone likelihood ratio family and $H : \lambda \leq \lambda_o$ against $\lambda > \lambda_o$, classical UMP tests exist; Meeden proved the existence of UMP Bayesian tests for the same case. Perhaps UMP Bayesian tests also exist, say for the same family and $H : \lambda \leq \lambda_1$ or $\lambda \geq \lambda_2$ against $\lambda_1 < \lambda < \lambda_2$. Because of lack of time it has not been possible to investigate the problems mentioned above.

LIST OF REFERENCES

[1] Lehmann, E.L. (1959) "Testing Statistical Hypotheses", Wiley, New York.

[2] Meeden, G. (1968) Doctoral Thesis, Chapter III. "Bayesian Hypotheses Testing as a Two-Set Prediction Problem", Iowa University.

[3] Meeden, G. (June 1970) "Best Tests for Testing Hypotheses About a Random Parameter with Unknown Distribution", The Annals of Mathematical Statistics, June 1970, 585-591.

[4] Neyman, J. (1962) "Two Breakthroughs in the Theory of Statistical Decision Making", Review of the International Statistical Institute, Vol. 30, 11-24.

[5] Robbins, H. (1950) "Asymptotically Subminimax Solutions of Compound Statistical Decision Problems", Second Berkeley Symposium, 131-140.

[6] Samuel, E. (1963) "An Empirical Bayes Approach to the Testing of Certain Parametric Hypotheses", The Annals of Mathematical Statistics, V34, 1370-1377.

[7] "Tables of the Bivariate Normal Distribution Function and Related Functions", National Bureau of Standards, Applied Maths. Series 50.